23

Pitch Perception

Exactitude is not Truth.

—Henri Matisse

23.1 Overview

The sense of *pitch* is more subtle than most of us think. It masquerades as objective but is, in fact, subjective. Pitch is a psychoacoustic phenomenon, a sensation, akin to hot, cold, or bitter, *synthesized* for our conscious minds courtesy of the ear-auditory cortex system. The raw data acquired by the outer, middle, and inner ear are passed first to the primary auditory cortex, which in turn produces a kind of *executive summary* of the sound, suitable for the busy conscious mind.

What would music be like without the summary sensations of loudness, pitch and timbre? Imagine instead being aware of all the individual partials and their relative strengths all the time. Hundreds of them could be competing for your attention. Music would be nothing like what we are familiar with. Far too much raw data would flood our consciousness.

Hearing a pitch does not mean that a partial or even a *tone* (a sound with partials related to the perceived pitch frequency) is present at the perceived pitch frequency. Pitch is perceived even for sounds that are *aperiodic*, such as a chime, whose partials are not evenly spaced; indeed, the perceived pitch does not coincide with any partial present.

The reader may object that on the contrary, pitch is quantitative. After all, some people have *perfect pitch*, meaning that they can name a note or hit the right key on the piano on the first try. Pitch, as we will see, is indeed quantitative in the sense that it is keyed to features in the autocorrelation of sounds. If you have perfect pitch, it is because you can match a signal with a prominent peak in the autocorrelation of the sound at, say, 0.0051 second with a key on the piano with a prominent peak in its autocorrelation at the same time. You will have hit the key G3. (Pitch is specified by non-perfect-pitch listeners by selecting a frequency of a pure partial that is judged to have the same pitch as the sound in question.)

Many sounds have no identifiable pitch. Other sounds may *seem* to have no pitch, but in fact a melody may emerge from a succession of similar sounds. Sets of wooden blocks have been fabricated going back to the nineteenth century to demonstrate this. If a block is dropped on a hard floor, the sound might be identifiable as containing frequencies in some range, but not possessing a pitch of any specific frequency. This impression might be reinforced by recording and Fourier analyzing the sound—it might show a range of seemingly unrelated frequencies. Yet if a number of similar carefully chosen blocks are dropped in succession, a familiar melody can force itself on the listener. Since melody is based on pitch, there must be a pitch present—at least when it is called to our attention. There is no correct answer to whether a pitch is present in the sound of a wood block, since the human subject is the ultimate authority, by definition. If the pitch was not heard, it was not present.

It is difficult to reason in a detached way about subjective sensations. If two people are coming from a different place in that debate, then something obvious to one person might be vehemently rejected by the other. This is a recipe for debate going around in circles, and indeed today you can find the same controversies that flared up in the mid-1800s.

23.2 Pitch Is Not Partial

The pitch of a 100 Hz pure sine tone is clearly 100 Hz; and that of a pure 200 Hz sine tone is of course 200 Hz (figure 23.1). In these cases, pitch and partial coincide in frequency. What is the pitch of both partials played together (figure 23.2)? It is not immediately clear that there will be a single pitch in the resulting complex tone. After all, there are two quite distinct partials present, well separated, and of equal power, so perhaps we register the presence of both, and report hearing the two partials present, one as important as the other. There would be nothing wrong with a hearing system that did this. But this is not what usually happens. The sensation of a single 100 Hz pitch usually prevails when both partials are played together. In this case of only two partials, one might become conscious of both at all times, especially if the partials had just been presented individually. But no one hears 10 separate "pitches" when 10 partials have significant strength.

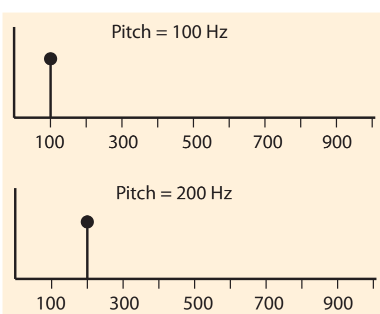


Figure 23.1
The pitch of these two partials is unambiguous.

Suppose now that we decrease the amplitude of the first 100 Hz partial (figure 23.3). Try this experiment in Jean-François Charles's MAX patch *Partials* or Paul Falstad's *Fourier*, but adjust the partial strengths with the sound off, since otherwise your attention will be drawn to them. At first, the pitch remains 100 Hz, and again we can still hear both partials with some concentration. If the 100 Hz partial is decreased toward zero amplitude, there will be only a pure 200 Hz sine tone remaining, so at some point the pitch has to switch to 200 Hz. This switch cannot be sudden: you are unlikely to hear a definite pitch of 100 Hz when the amplitude of the first partial is 5% of the 200 Hz amplitude, and then suddenly hear a definite pitch of 200 Hz when the amplitude of the first partial decreases to 4%. There must be a transition region, where both pitches are evident even if you are not trying to listen analytically. These sorts of "twilight zones" for pitch are commonplace.

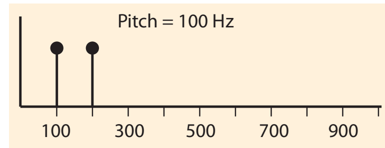


Figure 23.2
What is the pitch if both partials are played?

Discussions of pitch often mention only frequencies, and not amplitudes. Granted, the perceived pitch may not change much for wide variations of the amplitudes, but it will change if the amplitude changes become extreme enough. Any discussion of pitch should include mention of the amplitudes or power spectrum used—distrust any theory or opinion that does not. Incredibly, that will eliminate most of the theories out there!

The distinction between the perceptual, subjective nature of pitch, in contrast to the analytic, quantitative nature of partials, is reinforced by the *missing fundamental effect*, first brought to light by August Seebeck using sirens in the 1840s (see section 23.4). In the last example, we dropped the 100 Hz partial, leaving only a 200 Hz remaining partial and ending with a 200 Hz pitch. This is not surprising. However, the situation changes drastically if there are higher harmonics present initially, as in 100 Hz + 200 Hz + 300 Hz + 400 Hz + 500 Hz partials of equal amplitude. This complex tone has a pitch of 100 Hz. This time, when we drop the 100 Hz partial, leaving 200 Hz + 300 Hz + 400 Hz + 500 Hz partials, we hear a 100 Hz pitch, not 200 Hz as before. Another path to the same end is to add 300, 400, 500, ... Hz partials to a pure 200 Hz partial. Before the addition, the pitch was 200 Hz. All the new partials are higher in frequency than 200 Hz, yet a *lower* 100 Hz pitch develops again with no 100 Hz partial present. *The pitch is the frequency of a fundamental that is missing.*

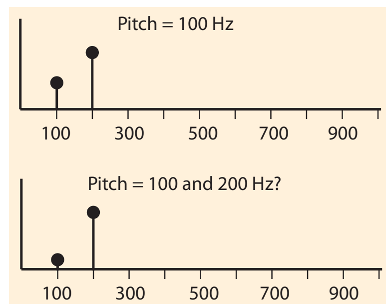


Figure 23.3
How does the pitch change as amplitudes change?

J. F. Schouten of the University of Eindhoven coined the term *residue pitch* in 1940 for the presence of a perceived pitch that is missing a partial at the same frequency, or even any nearby harmonics of that frequency. Giving a concept a name is extremely important, but apparently this came so late in the (by then) nearly 100-year-old controversy that people still trip over the issues. *Residue pitch* is a better term than the commonly used *missing fundamental effect*, since the latter phrase implies some sort of auditory illusion, which is not correct. The residue pitch is no more an illusion than is a yellow spot of light, when in fact the spot is made up of overlapping red and green beams. The result is indistinguishable to our eyes from a single yellow beam. This is not an illusion; this is the way our visual system is built and is supposed to act. So it is with the residue pitch: we hear a 100 Hz pitch, whether or not the 100 Hz partial is there. The executive summary sensation of pitch reports the period or more generally

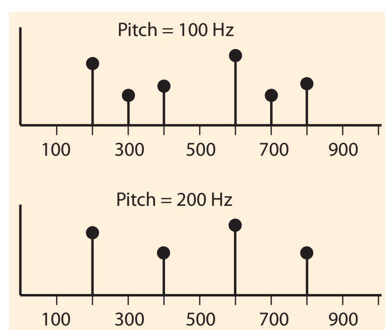


Figure 23.4
Two complex tones sharing some partials, but with very different pitch.

the presence of peaks in the autocorrelation of the sound, as we emphasize in the following.

MAX *Partials* or Falstad's *Fourier* can be used to check the residue pitch effect. Listen carefully for the presence of a 200 Hz sinusoidal partial with only partials 400, 600, 800, 1000, 1200, and 1400 present. You won't hear a 200 Hz *partial* at low sound intensities, yet a 200 Hz *pitch* will prevail. The 200 Hz partial is not there in the original sound, nor is it created for us by any part of the hearing system from the outer ear to the auditory cortex. What *is* created by that system is the sensation we call pitch. *Pitch is not partial.*

23.3 Pitch Is Not Periodicity

If pitch is not partial, the next line of defense might be that pitch is *periodicity*: the 100 + 200 Hz partials combined have an unambiguous 100 Hz periodicity—which might be used to "explain" why 100 Hz is the pitch heard. (See section 3.9 for a discussion of the periodicity of combinations of partials.) Early in his remarkable *On The Sensation of Tone as a Physiological Basis for the Theory of Music* (English edition, 1875), which is even today a foundation for psychoacoustics, Helmholtz states that pitch is periodicity. But consider this combination: 120, 220, 320, 420, 520, and 620 Hz in equal measure. The periodicity is 20 Hz, but the pitch is 104.6 Hz! Similar examples are examined quantitatively in sections 23.10 to 23.17.

If pitch is not determined by periodicity, perhaps then by autocorrelation? Now we are getting somewhere. The first major peak in the autocorrelation after the ever-present peak at time 0 in the preceding example occurs at 0.00956 second, which corresponds to 104.6 Hz. If sounds resolve into several distinct pitches, these will have corresponding peaks in the autocorrelation. *Pitch is associated with something more important than the presence or absence of a single partial: its tendency to repeat itself at given intervals. The generalization of this concept to sounds that are not strictly periodic is the autocorrelation function.* Periodic signals will always have peaks at multiples of the period that mirror the peak at time zero. The analogs of such periodicity are peaks in the autocorrelation of nonperiodic signals. We will explore the autocorrelation theme extensively in sections 23.10 to 23.17.

23.4 Pitched Battles

The subject of pitch perception heated up in the mid-nineteenth century with a debate between physicists Hermann von Helmholtz and Georg Ohm

Figure 23.5
The four nineteenth-century principals in the theory of pitch: Helmholtz, Ohm, Seebeck, and Koenig.

on one side and Rudolf Koenig and August Seebeck on the other (figure 23.5). They went to extreme lengths to try to achieve control of sound sources in order to settle ambiguities of human hearing. At some risk of oversimplification, we can state in a few words what the controversy is all about: Are human beings essentially walking Fourier analyzers?

The debate continues today, although it is slightly more subdued. In this chapter, we take a partly historical view, not only because of the fascinating personalities involved, but also because the old controversies are still in play today and still pose the appropriate questions. In so doing, readers will be empowered to form opinions on the controversies based on their own hearing, using modern apparatus that it is fair to say the principals mentioned earlier would have paid dearly for. It is far from the truth to say that everything is presently understood. Beautiful ideas that "ought" to be right, but unfortunately aren't, die hard. Some are still in the process of expiration.

Hermann von Helmholtz (1821–1894) was a towering figure in nineteenth-century physics. His theory of dissonance and musical harmony holds sway today; we discuss it at greater length in chapter 26. Helmholtz was perhaps the most renowned physicist of his day, brilliant and dominant in almost everything he did, including physiology, the theory of color vision, and the invention of the familiar ophthalmoscope used to examine the retina. Helmholtz was a talented musician and an expert in music theory. Thus he was in his element when dealing with both the physics and psychophysics of hearing and perception.

Georg Ohm (1789–1854) had a checkered academic career, partly university trained, partly self-taught, and later sometimes a high school teacher, sometimes a professor. Famously, he discovered the basic law of electrical resistance, which bears his name. Less famously but more important for this book, a second law also bearing his name pertains to the decomposition of arbitrary periodic sounds into sinusoidal partials.

August Seebeck (1805–1849) was a schoolmaster and physicist in Dresden, Germany, and son of physicist Thomas Seebeck, the discoverer of the thermoelectric effect whereby a voltage and electrical current is generated by a temperature gradient. Seebeck never held a university professorship (he was the head of the Technische Hochschule in Dresden), and he died at age 44. Using sirens, in 1841 he discovered the residue pitch effect wherein a pitch of frequency f is heard even though the only partials present are higher harmonics of f. Seebeck performed many other important experiments with sirens. Seebeck's scientific talents were not lost on Ohm and Helmholtz.

Rudolph Koenig (1832–1901) was an instrument builder par excellence. He spared almost no effort to create clean and concise experiments to test various aspects of human hearing, pitch perception, and phantom tone perception. See box 25.2 for more about this remarkable scientist-artisan.

23.5 The Siren

The siren played a key role in removing the umbilical cord that connected pure partials with pendular (sinusoidal) motion. It was a revelation that the extremely nonsinusoidal successive puffs of air pressure from a siren still produce upper partials that a trained ear can hear out as sinusoidal, ringing as true as if from a tuning fork. The physical fact that the partials were there but the disturbance creating the sound was nothing like a sinusoidally vibrating surface led Ohm to a new framework for understanding periodic sound in terms of Fourier's theorem.

Ohm realized that the source does not have to physically execute pendular vibrations in order to produce pendular, sinusoidal partials. The ear can't know what the source was physically doing, it hears only regular pulsations. The regular pulses can be mathematically Fourier analyzed into equally spaced sinusoidal partials. Despite this mathematical truth, it is still remarkable that the ear can perform such a Fourier decomposition. Ohm's contribution was twofold: not only is the decomposition of periodic tones into sinusoids always possible mathematically, but the sinusoids are really present, whether or not the sound was produced by pendular action. Ohm thus put this major misunderstanding (that the object producing the sound had to be manifestly sinusoidal in its vibration pattern) to rest by an application of Fourier's law. Strangely, he botched some important details (see section 23.8).

Jean-François Charles's MAX *Siren* (available on whyyouhearwhatyouhear.com) is a flexible siren simulator that can be used to reproduce many key experiments. Paul Falstad's *Fourier* or Charles's MAX *Partials* can be used to reveal the partials required to create a series of puffs.

23.6 Ohm's Law

The importance of the connection between sinusoidal waveforms and the pure tones—single partials—cannot be overemphasized; it is the one fixed boulder among many rolling stones in the field of pitch perception. It was Ohm who made this connection explicit. Helmholtz gave it legitimacy by making the connection a centerpiece of his work.

Ohm realized that only the sinusoid waveform yields the sensation of a colorless tone, a pure partial. Any embellishments to this sound taking the waveform away from a pure sinusoid requires higher harmonics (higher partials) to describe.

Ohm's advance was slow to diffuse its way into the fledgling world of psychoacoustics. According to the prevailing notion, the ear was supposed to be receiving souvenirs of motion in the object generating the sound. Ohm understood that, on the contrary, *any* periodic undulation could be decomposed into pendular (sinusoidal) components, and each partial would sound just as bright and clear whether some object vibrated exclusively at one frequency or at many frequencies at one time. Indeed, this was a straightforward application of Fourier's law from early in the century, but like so many other things in psychophysics, it is not always clear that nature has decided to follow the path of the mathematicians.

What about nonperiodic sounds, such as a chime? These too fall into the domain of the Ohm-Helmholtz laws—the partials in a chime tone are also pure sinusoids, except that they are not harmonically related.

Ohm's 1843 paper was unfortunately simultaneously pompous and muddled, as if to mask a measure of self-doubt. Helmholtz saw the significance of the paper more clearly than its author. Summing up what Ohm had done, Helmholtz said, "the proposition enunciated and defended by G. S. Ohm must be regarded as proven, viz. that *the human ear perceives pendular [sinusoidal] vibrations alone as simple tones*." This is true, and of unsurpassed importance in sound perception. But then Helmholtz reveals his own obsession with the human ear as a Fourier analyzer by continuing "*and resolves all other periodic motions of the air into a series of pendular vibrations, hearing the series of simple tones which correspond with these simple vibrations*" (emphasis is Helmholtz's).[1] It is true that all periodic motions of the air can be resolved *mathematically* into a series of pendular vibrations, but only the best, trained, or prompted ears can parse the sound into its partials,. and then only some of the partials. Even possessors of such ears normally listen holistically rather than performing the harder work of "hearing out" individual partials.

[1]From Helmholtz's *On the Sensation of Tone*, p. 56.

By 1937, Dayton C. Miller of Case School of Applied Science (now Case Western Reserve University), himself a formidable figure in the acoustics of his day, stated Ohm's law as follows:

> that all musical tones are periodic functions; that the ear perceives pendular [sinusoidal] vibrations alone, as simple tones; that all varieties of tone quality or tone color are due to particular combinations of a larger or smaller number of simple tones of commensurable frequencies; and that a complex musical tone or a composite mass of musical tones is capable of being analyzed into a sum of simple tones.[2]

It could not be stated better.

23.7 Seebeck's Mistake

Before Ohm's work, the sinusoid–pure partial connection had been blurry in several respects. Some observers thought that waveforms other than sinusoidal could also be perceived as pure partials, as long as they were periodic. August Seebeck fell into this trap, when trying to explain how it is that 100 Hz wins so handily in the simple "competition" for perceived pitch between 100 and 200 Hz (and higher) pure partials when they are both present. Seebeck supposed that somehow the 200 Hz component could add to the strength of the 100 Hz pure tone—that is, that the 100 Hz pure tone could be made louder by adding in some higher sinusoid of shorter but commensurate period. Seebeck arrived at this notion by throwing the presence of the period-reinforcing upper partials *onto the lowest partial*. He could not have meant this in a mathematical sense, since it violates Fourier's theorems, but rather in a physiologic sense. However, a strong sense of a 100 Hz pitch that accompanies the series 200, 300, 400, . . . Hz is not that of a 100 Hz fundamental sinusoidal partial. That sensation is absent, it cannot be "heard out," even though a 100 Hz *pitch* is definitely heard. Once again, *pitch is not partial*, a fact that both Helmholtz and Seebeck failed to see clearly.

23.8 Ohm's Blunder

The power of the (trained or prompted) ear to parse partials out of a tone induced both Ohm and Helmholtz to overplay the Fourier role in pitch

[2] D. Miller *The Science of Musical Sounds*, Macmillan, New York, 1926.

perception. This, in turn, probably caused Ohm to make a mathematical blunder. Helmholtz also could not resist the Fourier deconstruction of tone, and substituted his own idea of nonlinear effects to account for the pitch in the presence of missing partials.

Fourier's theorem allowed Ohm to write:

$$s(t) = a_1 \sin(2\pi f t + \phi_1) + a_2 \sin(4\pi f t + \phi_2) + a_3 \sin(6\pi f t + \phi_3) + \cdots .$$

Ohm knew that each of the terms on the righthand side corresponded to a *different* partial that could possibly be heard out by analytic listening. This much is true, but by itself it suggests a kind of democracy of partials, and doesn't explain our sense of pitch or, for example, the case of 100 Hz and 200 Hz, wherein a 100 Hz pitch is reported unless it is many times weaker than 200 Hz.

Ohm needed to explain why a siren with all its partials, many much stronger than the 100 Hz fundamental, should have a 100 Hz pitch if 100 holes were passing by the source of air per second. In fact, the lowest partial in a siren is usually quite weak. Conveniently for his prejudices, in the course of his lengthy and rather overly formal analysis Ohm made a mathematical blunder, which caused him to tremendously exaggerate the strength of the fundamental partial a_1 when the siren is emitting a sound with pitch f—that is, when f holes per second are being exposed to the air hose. Seebeck pointed out the mathematical error in a paper about his own experiments and theory concerning the operation of the siren. Apparently, Ohm was deeply embarrassed; his overly formal paper seemed hollow in the face of such a mistake. Ohm got out of the field of acoustics altogether, but it turned out he had underestimated his own contributions.

After putting Fourier's theorem in proper context and connecting it with our ability to hear partials individually, Ohm and Helmholtz focused too much on the ear's analytic Fourier analysis capabilities, never assigning a role to any holistic synthesis. When it came to explaining pitch, Ohm couldn't let go of the idea that what we hear is a collection of partials, so no pitch could be heard unless there was a partial present at that frequency.

23.9 Helmholtz Falls Short

Helmholtz didn't do much better, although this point is still controversial. Helmholtz knew that Ohm was correct about the *principle* of Fourier decomposition of sound into pure partials, although he too must have winced at Ohm's mathematical blunder. He would also have been frustrated by Seebeck's confusion about the strength of the lowest partial depending on upper partials, exactly the point that Ohm had cleared up. But Helmholtz needed some other way to explain why a pitch of 100 Hz needed little or no power at 100 Hz in the tone. Here, he would soon make his own gaffs.

Helmholtz began by using bottles as resonators to detect partials, but Rudolf Koenig optimized them in brass (see figure 13.3), making a cavity with a large opening with a very short neck on one side and a small nipple on the other for insertion into the ear. These are the famous *Helmholtz resonators*, and as with other Helmholtz inventions, they were turned into something of an art form by Koenig, prized by museums of scientific instruments today. These resonators are relatively high Q and respond only to a very narrow range of frequencies. With them, Helmholtz could easily verify the presence or absence of an objective partial at a perceived pitch, since it would be so much enhanced if present.

The principle is not near-field capture (NFC), since the source may not be close by, nor is the source made louder, except inside the resonator. The idea is to set up a Helmholtz resonance in the usual way and then *listen to what is happening on the inside of the resonator*. The sound is *much* louder there, but we normally cannot hear it.[3] However, if a small nipple protrudes out the back of the resonator, tightly sealed in the ear canal, the nipple and short air cavity leading to the tympanum become part of the inside cavity. The tympanum is subjected to the full SPL inside the resonator, greatly enhancing any partial present at the resonator's frequency. This is why Helmholtz resonators work so well, a fact seemingly almost forgotten since Koenig's day.

Helmholtz knew that a pitch at frequency f could be heard with very weak or absent partials at f, since his resonators failed to find them in some circumstances. His theories of combination tones, to be taken up in chapter 25, appeals to mechanical nonlinear interactions in the ear to create the fundamental partial missing in the arriving signal. This idea, which once again confused pitch and partial, does not stand up to scrutiny.

This error by such a great scientist is surprising, and reflects how even the best scientists struggle with objectivity when the subject of their experiments is themselves. In his book *On the Sensation of Tone*, Helmholtz reveals just how comfortable he is with problems of perception, freely acknowledging of the role of synthetic listening:

> We . . . become aware that two different kinds or grades must be distinguished in our becoming conscious of a sensation. The lower grade of this consciousness, is that where the influence of the sensation in question makes itself felt only in the conceptions we form of external things and processes, and assists in determining them. This can take place without our needing or indeed being able to ascertain to what particular part of our sensations we owe this or that relation of our perceptions. In this case we will say that the impression of the sensation in question is perceived synthetically. The second and

[3]Except inside a car traveling down the highway with one window open. However, our hearing is not sensitive to sound at the frequency produced, but the SPL is so high that we can feel it!

higher grade is when we immediately distinguish the sensation in question as an existing of the sum of the sensations excited in us. We will say then that the sensation is perceived analytically. The two cases must be carefully distinguished from each other.

It is all the more surprising after this eloquent summary that Helmholtz did not assign a synthetic role to the sensation of pitch. Moreover, Helmholtz's theories of "tokens" or "signs" were part of a sophisticated understanding of epistemology that could have cleared this up, but instead he followed Ohm, apparently failing to recognize pitch as one of his tokens! Helmholtz dismissed pitch as periodicity, sidestepping issues of missing fundamentals and nonperiodic tones. Later, in discussing combination tones, Helmholtz wrongly attributes the perceived pitch to a partial *created in the ear* by nonlinear interactions as we mentioned earlier. (This may actually happen for very loud tones, however.) Helmholtz apparently thought that hearing a tone or pitch of f meant a partial at f had to be present, but his discussion is ambiguous because he is imprecise about the presence of pure partials in the perceived tone. This absolutely key point was muddled up in *On the Sensation of Tone*, for, as his translator, John Ellis, said, "Even Prof. Helmholtz himself has not succeeded in using his word *Ton* consistently for a simple tone only" (that is, a simple partial). This is the one thing he should have made crystal clear, but he repeatedly fails to do so: Does one always hear a sinusoidal partial oscillating at any given perceived pitch? The answer is, clearly, no, but Helmholtz never quite framed the question this way.

The key point is not to confuse pitch with presence of a partial at the frequency of the perceived pitch. Create the tone $200 + 400 + 600 + 800$ Hz, in *Fourier* or *Partials*, and then raise and lower the amplitude of the 200 Hz partial, all the way to zero. Concentrate on what a 200 Hz partial sounds like, and try to hear it when it is at zero amplitude. You won't hear that partial, but you will hear the 200 Hz pitch.

Three of four excellent scientists made serious but different mistakes when trying to explain the dominance of the lowest "root" fundamental in pitch perception, whether or not a partial is actually present at the root frequency. Again, the fact that such talented people made mistakes testifies to the pitfalls associated with trying to be objective about one's own subjective sensations.

23.10 A Dramatic Residue Pitch Effect

To drive an important point home, consider figure 23.6 (and the sound file 50 Hz Missing Fund, available on whyyouhearwhatyouhear.com). This track begins with a 200 Hz periodic tone with partials at 200, 400, and

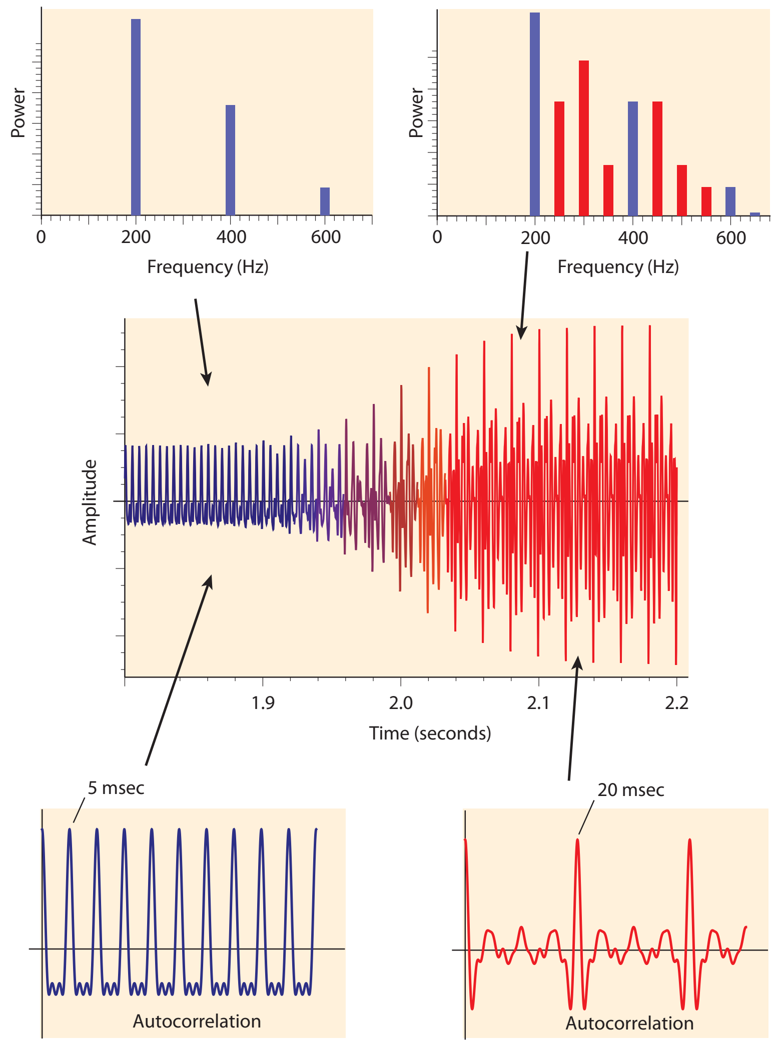


Figure 23.6

The sound trace for the transition region for the example audio file, 50HzMissingFund. Up to about 1.9 seconds, the sound is an ordinary complex 200 Hz tone with three partials, but after 2.1 seconds it has a partial every 50 Hz starting at 200 Hz and ending at 650 Hz. This latter progression has a 50 Hz frequency—four times smaller than the 200 Hz frequency at the beginning. This longer period can be seen in the trace after 2 seconds. A strong sensation of 50 Hz sound emerges as the new partials come in, but there is no 50, 100, or 150 Hz component at all. The autocorrelation functions (bottom) reveal the transition from a 200 Hz pitch to a 50 Hz pitch.

600 Hz. The power spectrum is shown at the upper left in figure 23.6; the corresponding sound trace is shown in the first part of the middle panel. The autocorrelation is shown at the lower left. The pitch heard is 200 Hz, and the autocorrelation has a prominent peak at $1/200 = 0.005$ s. Starting just before 2 seconds into the file, partials at 250, 300, 350, 450, 500, and 550 Hz are added. *Despite the fact that all of these partials are higher in frequency than the original perceived pitch and higher than the lowest partial originally present, the pitch drops by two octaves to 50 Hz!* There is no 50, 100, or 150 Hz component at all. The $T = 0.02$ second periodicity, corresponding to 50 Hz, is clearly seen after the 2-second mark

in the sound trace. The GCD of 200, 250, 300, . . . is of course 50 Hz, which as we discovered in section 3.9 is the period of the combination of the preceding partials. The autocorrelation now has its first prominent peak at $1/50 = 0.02$ s, as seen in the lower right.

Truth or Illusion?

The acoustics group at the University of New South Wales dubs the residue pitch effect an "auditory illusion," which is another way of saying that pitch is not really there. This is compatible with the idea that pitch is a sensation like hot or cold, but perhaps the word *illusion* is too strong, because it is by design that we process pitch the way we do.

When confronted with the strong dominance of the fundamental over higher partials in musical tones, Ohm also referred to auditory illusions or tricks that the mind was playing. He viewed this as some kind of an anomaly, rather than a necessity or at least a preference of the human mind.

We prefer the terms *executive summary* or *token* of reality rather than *illusion*, since many illusions, especially visual ones, are unexpected and sometimes just plain weird side effects of the way our sensory systems work. The sensation of pitch is not a weird side effect. It serve a purpose. If something is vibrating at 100 Hz, we are much better off hearing a 100 Hz pitch, which is telling us the truth: the object is vibrating at 100 Hz. The fourth overtone partial at 400 Hz might be the loudest frequency arriving at our ears, and 100 Hz may be absent, but why would we want to be distracted by that? The pattern of partial strengths is cast into the sensation of timbre.

Small speakers in a laptop are very poor at creating low-frequency sound. If an object producing the sound is much smaller than the typical wavelengths of the sound produced, the pressure (force) and the acceleration at the surface of the object are nearly in phase, as explained in section 7.11. Once the object is appreciable in size compared to the wavelength, it is possible for the force and velocity to be more nearly in phase, greatly enhancing the work done on the air by the vibrating object, and therefore its loudness. Small laptop speakers can produce only a very weak tone if driven sinusoidally at 100 Hz (wavelength about 3.5 meters, much larger than the speakers). However, because of the residue pitch effect, the same speaker producing a 100 Hz *complex* tone gives rise to the strong sensation of a 100 Hz pitch, in spite of the near total absence of a 100 Hz partial, and a very weak 200 Hz partial.

23.11 Autocorrelation and Pitch

The dramatic change from a 200 Hz pitch to a 50 Hz pitch after adding partials no lower than 250 Hz was accompanied by a shift in the first large peak in the autocorrelation from 5 ms to 20 ms—that is, the inverse of 200 Hz and 50 Hz, respectively. The 50 Hz pitch was heard in the absence of the first *three* partials—namely, 50, 100, and 150 Hz.

The idea that autocorrelation is what determines pitch came rather late, only in 1951, suggested by J.C.R. Licklider. The notion seems to have had a rather lukewarm reception in the literature ever since, yet autocorrelation is what pitch estimators use in many sound analysis programs, such as Praat and Audacity. Physiologically, it is not clear whether autocorrelation is literally computed in a neural circuit or merely strongly related to whatever is. The autocorrelation idea is an example of a temporal theory of pitch perception. We will expand on how autocorrelation may be used to determine pitch shortly, but it works so well in so many circumstances that it seems safe to say this: *Beware of any theory of pitch perception that entirely leaves out autocorrelation.*

Autocorrelation was defined in chapter 4. It can be constructed from the power spectrum, and is therefore equivalent to it. A peak in the autocorrelation function at a time τ means that the function *tends* to be similar to itself at times t and $t + \tau$, for all t. Unlike periodic sounds, which are strictly correlated with themselves (doing the same thing at the same time intervals forever), a less than perfect correlation peak (peak height less than one) only implies a tendency to mimic what came a time τ before.

The autocorrelation predicts the residue pitch effect. In this case, the peaks in the autocorrelation function reflect what we already knew from the periodicity of the sound. Strict periodicity is reflected in autocorrelation peaks that are as prominent as the first peak at time zero. The ultimate test of the autocorrelation idea involves perceived pitches of nonperiodic sounds.

An early, tall, isolated peak in the autocorrelation function will determine a perceived pitch, as the inverse of the time of the peak. Clearly, given some wiggly autocorrelation function, notions of *early*, *tall*, and *isolated* are qualitative at best. However, this is just as it should be. Pitch is itself qualitative—its sensation can be weak or strong, there can be more than one pitch present, and attention can be focused on each pitch separately. (Just as we can hear out partials, we can also hear out separate notes—separate collections of partials—even though this ability may require contextual cues.)

23.12 A Simple Formula for Pitch

A simple, approximate formula for the pitch seems to work very well in a reasonable range of circumstances. The idea is to find a good approximation to the time of early, tall autocorrelation peaks, given the set of amplitudes and frequencies as input. At the top of this peak, the slope of the autocorrelation function is zero; an approximation is developed for the time of such a peak, and then its inverse gives the frequency.[4] Given a set of frequencies f_n and amplitudes a_n (power $p_n = a_n^2$), the virtual pitch $\bar{f}$ that will be heard is given by

$$\bar{f} \approx \frac{\sum_n a_n^2 f_n^2}{\sum_n a_n^2 N_n f_n} = \frac{\sum_n p_n f_n^2}{\sum_n p_n N_n f_n}, \tag{23.1}$$

where N_n is an integer depending on f_n: $N_n = [f_n/\bar{f}]$, where $[\ldots]$ is the integer nearest to the quantity inside the brackets—for example, [5.23] = 5; [4.9]=5. This definition is slightly circular in that $\bar{f}$ depends on the integers N_n, which itself depends on $\bar{f}$, but in practice a self-consistent set of integers can usually be found.

As a test of the formula, we try the frequencies 820, 1020, and 1220 Hz of equal amplitude. The GCD of 820, 1020, and 1220 is 20, right at the threshold of hearing. This pitch seems an unlikely perceptual result of combining these much higher frequencies. In his book *The Science of Musical Sound*, John R. Pierce cites this case as an interesting example and reports that the perceived pitch is 204 Hz; formula 23.1 using $N_1 = 4$, $N_2 = 5$, $N_3 = 6$ gives 203.9 for amplitudes $a_1, \ldots = 1, 1, 1$. (Pierce did not report the amplitudes, but, by experimenting with the formula, it is found that the frequency is only mildly sensitive to them within reasonable limits.) Figure 23.7 makes the situation clear. The autocorrelation function is shown as a thick black line, and the individual cosine terms contributing to the autocorrelation are shown in color. The small numerals near the peaks of the cosines count the number of full oscillations starting at time equal to zero. Near time $t = 0.0049$, the 820 Hz frequency has oscillated four times, the 1020 Hz frequency five times, and the 1220 Hz frequency six times; thus $N_1 = 4$, $N_2 = 5$, $N_3 = 6$. A large peak rises at $t = 0.004904$, since all three cosines return to 1 near this time, although not exactly at the same time. The corresponding frequency is $f = 203.9 = 1/0.004904$ Hz. In spite of earlier recurrences (peaks), which would correspond to higher frequency pitches, this later recurrence is much stronger and dominates our sense of pitch. Precise measurement of the recurrence time from the autocorrelation function and formula 23.1 both give $f = 203.9$ Hz. By plotting the autocorrelation function for a much longer time (figure 23.8), we can easily see why 20 Hz is not the perceived pitch. There are many strong recurrences reached before 50 ms, which is the time corresponding to 20 Hz, and even

[4]In calculus language, we take the derivative of the autocorrelation and set it equal to zero, $dc(\tau)/d\tau = 0$, and search near the recurrence we're looking for. Using the approximation $\sin(x) \approx x$, valid for small x, we get formula 23.1. This formula was first used in 1982 in the context of molecular spectroscopy, by J. Zink and the author.

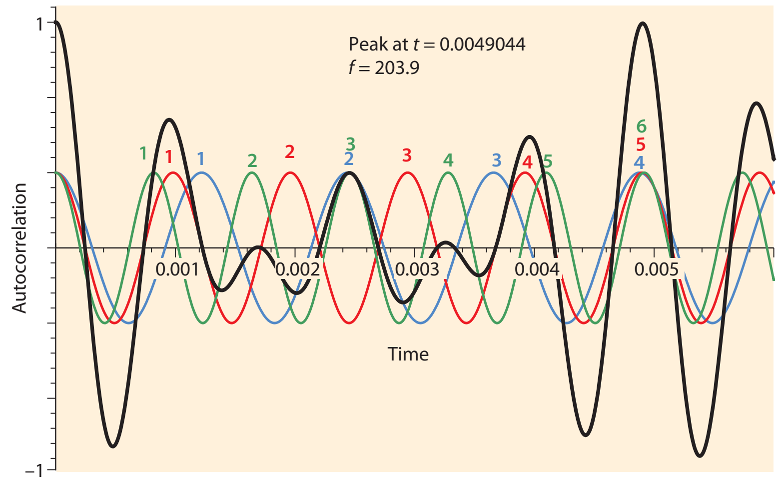


Figure 23.7
Autocorrelation function (black curve) analyzed for the perceived pitch corresponding to frequencies 820, 1020, and 1220 Hz with equal amplitudes. The autocorrelation function is the sum of the cosines shown in color.

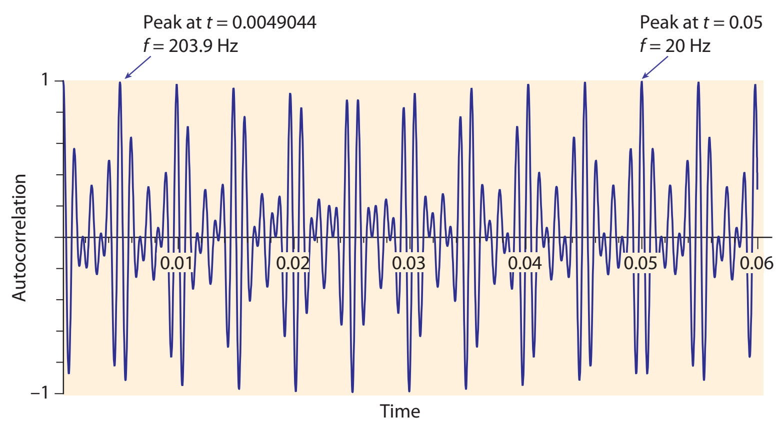


Figure 23.8
Longer time autocorrelation function analyzed for the perceived pitch corresponding to frequencies 820, 1020, and 1220 Hz with equal amplitudes. The autocorrelation function has a perfect recurrence at 50 ms, but it is only slightly higher than the many that have come before.

though the strongest one (by a slight amount) occurs then, apparently the earlier peaks have drawn our attention to higher pitches.

A similar problem was considered by R. Plomp in 2001 (R. Plomp, *The Intelligent Ear*), using the frequencies 850, 1050, 1250, 1450, and 1650 Hz, which have a GCD of 50 Hz. Plomp reported that people perceive "about 210" Hz. The autocorrelation function peak suggests 209.2; and formula 23.1 using $N_1 = 4, \ldots, N_5 = 8$ gives 209.13 for amplitudes $a_1, \ldots = 2, 2, 1, 1, 1$. Plomp did not seem to favor the autocorrelation idea; he advanced several other explanations for the apparent frequency shift.

We can check the autocorrelation formula against the class of examples suggested by Rausch and Plomp,[5] who plotted the residue pitch (they

[5] R. A. Rausch and R. Plomp, "The Perception of Musical Tones," in *The Psychology of Music*, ed. D. Deutsch, Academic Press, New York, 1982.

October 9, 2012 Time: 09:34am chapter23.tex

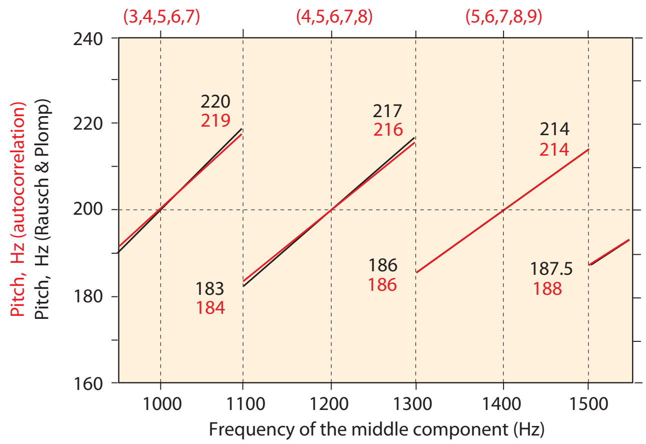


Figure 23.9
Residue pitch against c, for the series $200 + c$, $400 + c$, $600 + c$, $800 + c$, $1000 + c$, for c on the interval (350, 950). Black lines and numbers: Results of Rausch and Plomp. Red lines and numbers: Autocorrelation results (both numerical and from formula 23.1—they are very close to each other). The appropriate integers N_n to use in formula 23.1 are shown in red at the top.

called it the *low pitch*) against c, for the series $200 + c$, $400 + c$, $600 + c$, $800 + c$, $1000 + c$, for c on the interval (350, 950). Their results, based on experiments with volunteer subjects, are shown in figure 23.9, taken from the article in Deutsch's book, *The Psychology of Music*, along with our autocorrelation results. As c is increased, the appropriate integers N_i change, and are given at the top of the figure. The pitch obtained by autocorrelation (either by numerical peak finding, or from our simple formula 23.1—the results in this case differ by less than 1 Hz over the whole range) are shown in red. It is seen that the autocorrelation gives an essentially perfect estimate of the perceived pitch. At 1100, 1300, and 1500 Hz, there is an abrupt discontinuity in pitch, and at those frequencies the dominant pitch is indeed ambiguous.

The residue pitch formula estimates the time of maxima in the autocorrelation function, and therefore the corresponding pitch frequency. We can change which peak is being estimated by adjusting the N_i. However, this does not say which pitch dominates in *marquee effect* cases (see the following).

23.13 Examples: Autocorrelation and Pitch

A periodic signal with period T is perfectly correlated with itself at multiples of the period: whatever its value at time t, it is duty bound to be the exactly same a time T later, or any multiple of that time later. Likewise, $c(nT) = \langle y(t)y(t + nT)\rangle = \langle y(t)^2\rangle = c(0)$ is perfectly correlated, where n is an integer. So, for a periodic signal, we expect the autocorrelation $c(\tau)$ to be big at $\tau = 0$, and the same value again at times $\tau = nT$. In terms of our

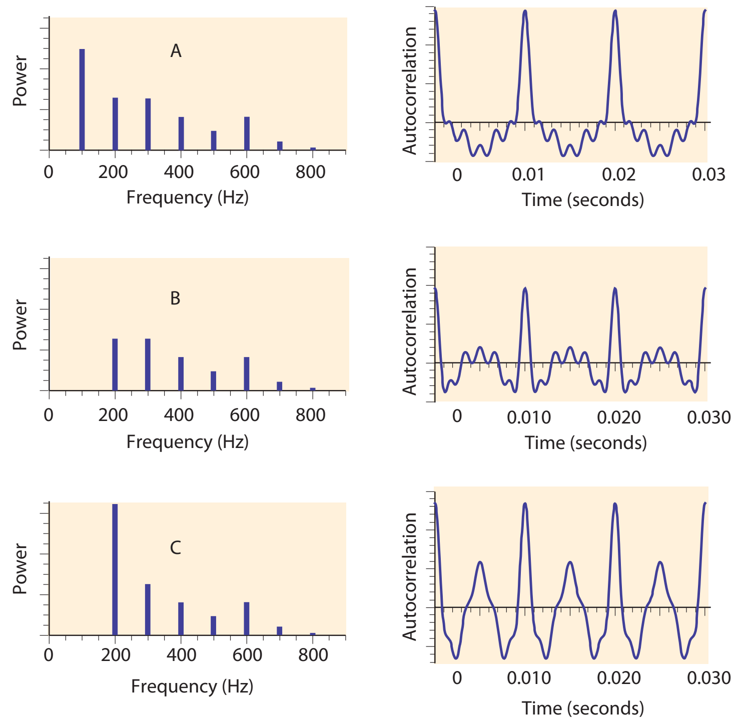


Figure 23.10
A: A "normal" power spectrum based on a 100 Hz fundamental (left), and its autocorrelation (right). Notice the periodic revival at $T = 0.01, 0.02, \ldots$ second. B: With the fundamental removed (left), the autocorrelation (right) is still periodic, with the earliest strong revival again at $\tau = 0.01$ second. C: By increasing the strength of the 200 Hz second partial, a revival at $\tau = 0.005$ second begins to form. Eventually, this becomes the dominant early revival, and our ear-brain system will switch over to hearing a 200 Hz tone, rather that the 100 Hz residue pitch.

formula,

$$c(\tau) = a_1^2 \cos(2\pi f \tau) + a_2^2 \cos(4\pi f \tau) + a_3^2 \cos(6\pi f \tau) + \cdots . \quad (23.2)$$

This is clearly periodic with period $T = 1/f$ as expected. Since $\cos(0) = \cos(2n\pi) = 1$, the correlation is large and positive at $\tau = 0$ and every period T thereafter. We show $c(\tau)$ for a typical periodic tone in figure 23.10A. Figure 23.10B shows the power spectrum and autocorrelation in A with the fundamental at 100 Hz removed. Last, in C, we see the spectrum and autocorrelation with the fundamental at 100 Hz still removed, but with the second partial at 200 Hz boosted. Now we notice a stronger, but still not dominant, revival at $\tau = 0.005$, although the periodicity is still $T = 0.01$ second. If the 200 Hz partial towers over all others, we will start to hear a 200 Hz pitch.

What does our pitch formula predict in the case of the residue pitch effect—for example, (200, 300, 400, ...)—as the amplitudes of the partials are varied? The GCD of (200, 300, 400, ...) is 100, and the period $T = 1/100$, the same as if the fundamental were present. The period corresponds to a frequency of 100 Hz, but the lowest frequency present is

200 Hz. We have

$$\bar{f} \approx \frac{\sum_n a_n^2 f_n^2}{\sum_n a_n^2 N_n f_n} = \frac{\sum_{n=2} a_n^2 n^2 f_0^2}{\sum_{n=2} a_n^2 n f_0 \cdot n} = f_0 \frac{\sum_{n=2} a_n^2 n^2}{\sum_{n=2} a_n^2 n^2} = f_0, \tag{23.3}$$

that is, it predicts the residue pitch heard is f_0.

Consider the series 300, 500, 700, ... Hz. This has a residue pitch of 100 Hz. Pierce[6] claims that successive odd harmonics of a missing fundamental do not produce the residue. Backus[7] on the contrary, says 300, 500, 700 will have a pitch of 100 Hz, which is just such a case. *Fourier* or *Partials* may be used to to arrive at your own resolution of these conflicting claims. The autocorrelation has a strong peak at 0.01 second, corresponding to the residue 100 Hz.

Rausch and Plomp have suggested several interesting examples:

- If we take partials at 850, 1050, 1250, 1450, and 1650 Hz, all of equal intensity (it is suggested that you try this in *Partials*), there is a strong autocorrelation peak corresponding to 207.90 Hz. This is indeed the perceived residue pitch. Rausch and Plomp report 208.3 Hz, and formula 23.1 gives 207.91, insignificantly different.
- Rausch and Plomp have claimed a *dominance region* in the frequency range 500 to 2000 Hz, suggesting that frequencies in this range are more dominant in determining the pitch than frequencies that are lower or higher. As an example, they give the frequencies 204, 408, 612, 800, 1000, and 1200 Hz. The first three partials alone give a pitch of 204 Hz. As an example of the dominance idea, Rausch and Plomp then report that the result of all six partials is 200 Hz, noting that the last three partials alone give this residue, and are within the dominance region. However, without assuming any kind of special dominance, the autocorrelation peak formula 23.1 gives 200.62 Hz with equal amplitudes for all partials, a near perfect perceptual match. This actually shows that the dominance idea has some merit, because higher frequencies, being made up of shorter wavelengths, make sharper peaks; adding sharp peaks to broader peaks coming from lower frequencies readily shifts the new combined peak to be near the sharper peak of the two.

Sounds may have more than one recognizable pitch, as in a musical chord on a piano or four voices in harmony. However, here we have to be very careful to acknowledge musical context as part of our ability to parse separate notes with different pitches from a sound. Recording a single piano chord and later rather clinically playing a sound bite back to

[6] John R. Pierce, *The Science of Musical Sound*, rev. ed., Freeman, New York, 1992, p. 95.

[7] John Backus, *The Acoustical Foundations of Music*, 2nd ed., Norton, New York, 1977, p. 130.

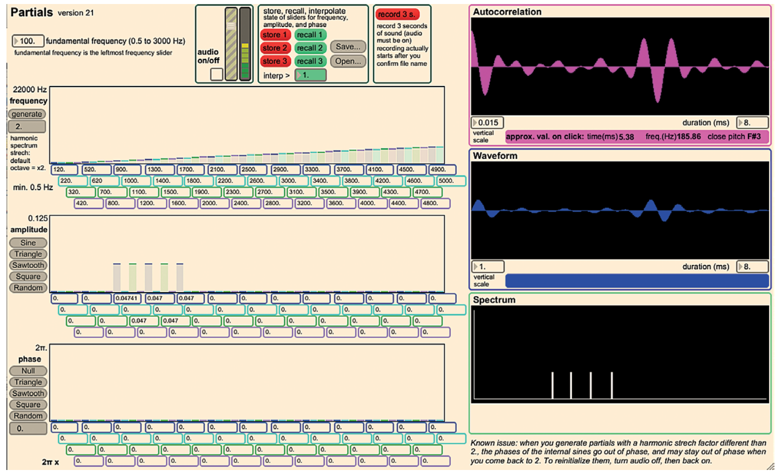


Figure 23.11
D. Deutsch example: 900, 1100, 1300, 1500, and 1700 Hz, run in the MAX patch *Partials*. Two autocorrelation peaks, corresponding to about 216 and 186 Hz, are revealed.

a listener, out of context, with no attack or finish, could result in a quite different impression of the sound, compared to listening to the same chord during a piano recital.

Consider the case 900, 1100, 1300, 1500, and 1700 Hz, suggested by Deutsch. She states that it is ambiguous, either 216.6 or 185.9 based on pattern matching (one of the theories of pitch that we will not treat) with a harmonic series. Indeed, the autocorrelation gives healthy peaks corresponding to both 215.78 and 186.4 (see figure 23.11).

Recent neurophysiological research has shown that the residue pitch is established in the auditory cortex within 1/10 of a second of the onset of the sound.

23.14 Seebeck's Pitch Experiments

A clever experiment by August Seebeck reveals much about autocorrelation and our built-in pitch algorithms. It is related to the 100 + 200 Hz partials of varying relative strength introduced earlier, but it deals with complex tones and springs from a physical sound source—a siren. Seebeck's improved siren, shown earlier in figure 7.19, had eight rows of holes and 10 adjustable compressed air tubes that could be placed on different rows and adjusted so as to cause the holes to be exposed at various phases relative to the other rows.

Seebeck drilled 60 holes in one circular row, but rather than space them evenly he offset them slightly in pairs, alternating the angle from one hole to the next: first 5 degrees, then 7 degrees, then 5 again, and so on. That is, there were pairs of holes 5 degrees apart separated by slightly larger

October 9, 2012 Time: 09:34am chapter23.tex

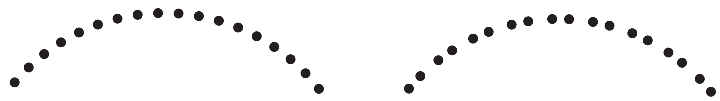


Figure 23.12

(Left) Siren holes spaced 6 degrees apart. (Right) Siren holes spaced alternately 5 degrees and 7 degrees apart. The period becomes twice what it was with all the holes 6 degrees apart.

7-degree gaps between holes on adjacent pairs. If all 60 holes had been instead each 6 degrees apart—evenly spaced—the result of rotating the disk five times per second would clearly be a 300 Hz complex tone with a 300 Hz pitch. The uneven set of holes corresponds to taking every other hole in the even set and rotating it by one degree. If the disk is spun again at 5 Hz, 300 holes pass by the air source per second, the same number as when the holes were exactly evenly spaced. It seems that the pitch should again be 300 Hz. Instead, the pitch drops an octave, to 150 Hz (figure 23.12).

MAX *Siren* (see figure 7.18) can be used to reproduce the experiment. Create two rows of holes, 30 each, with zero phase offset. At five revolutions per second, a 150 Hz pitch is very strong. The holes are sounding in pairs at exactly the same time, so the periodicity is 150 Hz. Now, using the phase tool, offset one row of holes by 6 degrees ($0.016666 \times 2\pi$ radians; type 0.016666 in the box). This setting causes the second set of 30 holes to sound exactly halfway between those of the first set, making a 300 Hz periodicity with the disk revolving at 5 Hz. The 6-degree offset setting is equivalent to 60 evenly spaced holes in the same row. (We can be sure that the second row of holes is providing exactly the same pressure profile at each hole as this first row is, because the sound is being produced electronically. Seebeck needed to place all the holes in the same row, to ensure that the holes are all given exactly the same air pressure.)

Up to now, we have established that with zero offset, the pitch is 150 Hz, and with a 6-degree offset, the pitch is 300 Hz. What happens in between? Do we hear both pitches in varying degrees? A 5-degree offset of the second set of 30 holes ($0.013888 \times 2\pi$ radians) was Seebeck's choice. Perfect 300 Hz periodicity happens only at 6 degrees; 5 degrees is not quite periodic at 300 Hz and strictly periodic at 150 Hz. Even though 300 holes are still passing by the air source per second, the perceived pitch drops an octave, to a strong 150 Hz with a 5-degree offset.

Since we don't have to laboriously drill holes, it is tempting to see what happens as we change the offset from 5 degrees back toward 6 degrees. To the author, both pitches are apparent at 5.5 ($0.015277 \times 2\pi$ radians) degrees offset, and the 300 Hz pitch becomes perhaps slightly dominant at 5.75 degrees. That is, not until the holes are almost perfectly evenly spaced does the pitch finally start to switch to the higher frequency.

The data are summarized in figure 23.13, which shows the autocorrelation graphs and the power spectra for 6, 5.75, 5.5. 5.0, and 0 degrees offset of the second set of 30 holes.

October 9, 2012 Time: 09:34am chapter23.tex

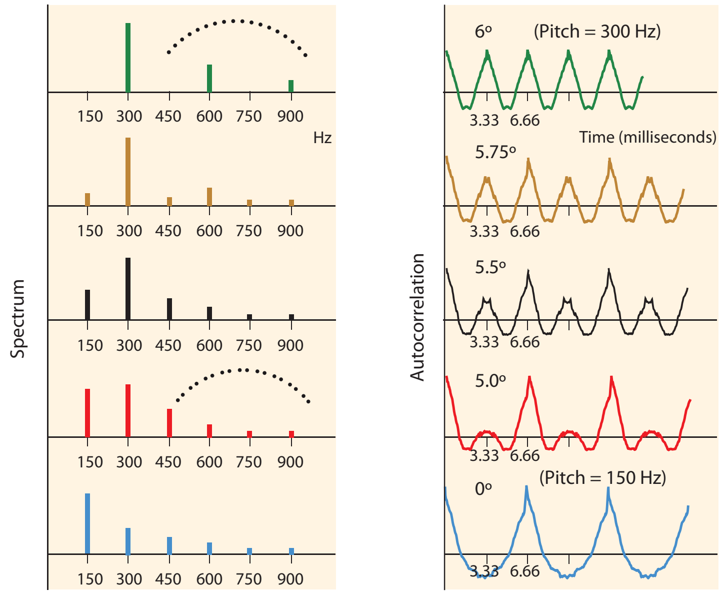


Figure 23.13
Autocorrelation and spectrum plots for the Seebeck siren experiment, wherein exact periodicity of the holes is slightly broken in favor of pairs. Seebeck used 60 holes, with each hole spaced by 5, 7, 5, 7, . . . degrees from its neighbor, where 6, 6, 6, 6, . . . would be even spacing of 60 holes.

The Marquee Effect

These experiments and the data in figure 23.13 give clues as to our built-in neural algorithms for determining pitch. The first and last cases are unambiguous, with the first large autocorrelation peaks at 3.333 and 6.666 milliseconds, corresponding to 300 and 150 Hz, respectively. The spectra reflect this periodicity. The middle cases are ambiguous and instructive. The game being played is to decide which autocorrelation peak determines pitch. It is possible to have a near tie, in which case we will hear two distinct pitches. The rules seem to be

- *Key point*: Earlier peaks are favored, taller peaks are favored, sharper peaks are favored.

There are limitations. For example, peaks may arrive too late to control our sense of pitch, even if they are tall. These rules can lead to a tie between an earlier, slightly smaller peak, giving a high-frequency pitch, and a later, taller peak, giving a lower pitch. It is a little like two stars who both want top billing on a movie marquee. Someone gets first billing, don't they? Not necessarily. Who is first in the marquee shown in figure 23.14? You may have a definite answer in this case, but not everyone

Figure 23.14
Who's got first billing? This is analogous to one form of the octave ambiguity problem.

will agree. The position of the names can be manipulated until they have equal billing for a given person. All this is in good analogy with pitch—people will switch over to hearing both pitches as equally important at different points in the competition between first, sharpest, and tallest. We dub the "earlier peaks are favored, taller peaks are favored, sharper peaks are favored" the *marquee effect*.

The marquee effect model of pitch perception serves as a rough guide to the pitch(es) we perceive. It is a very useful exercise to set up MAX *Partials* with 12 or 15 partials and manipulate them, watching the effect on the autocorrelation function and listening to the pitch. The pitches heard are definitely context dependent; you hear different things depending on whether you leave the sound on while switching partial strengths, and so on.

23.15 Shepard Tones

One of the most famous auditory demonstrations is called the *Shepard-Risset tones*, or *Shepard tones*, after the inventors. By a very clever choice of the amplitudes and frequencies of the partials, Shepard tones present a rising pitch from one semitone to the next. But after 12 rising semitones, the pitch winds up where it started! This feat is frequently compared to the impossible Penrose stairs, a 2D drawing of a 3D staircase invented by the physicist Roger Penrose and his father (figure 23.15). Every step is up (or every one down in the other direction), and yet one returns to the starting place. The illusion springs from an ambiguity of a two-dimensional rendering of what is in reality a three-dimensional object. The Shepard-Risset tone illusion stems ultimately from a pitch ambiguity, and we shall analyze it in several ways, as it is quite revealing.

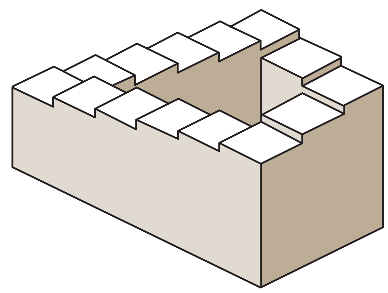


Figure 23.15
An impossible (in 3D) Penrose staircase.

It is not difficult to explain how Shepard tones actually work, yet this is seldom done. An equal-tempered scale climbs frequency as factors of $2^{1/12}$ per semitone. Normally, each note would have all the partials above it as integer multiples of the base frequency. However, Shepard used only a subset of these, those that are powers of 2, i.e., 2^n, $n = 0, 1, 2, \ldots$ above the first partial. The other partials are given no amplitude. The frequencies used in a complete octave climb up 12 semitones are then

$$f_{m,n} = 2^{m/12} 2^n f_0; \quad m = 0, 1, \ldots, 11; \quad n = 0, 1, 2, \ldots. \tag{23.4}$$

If $m = 0$, the first step in the sequencing of seeming rising pitch, the frequencies are $f_0, 2f_0, 4f_0, 8f_0, \ldots$. If $m = 12$, the tone is back to exactly where it started; the frequencies are $2f_0, 4f_0, 8f_0, \ldots$, almost the same; except missing f_0. Here is where the amplitude management comes in: an envelope is used that modulates the partials according to a fixed function or shape. The amplitude of a partial depends strictly on the frequency of

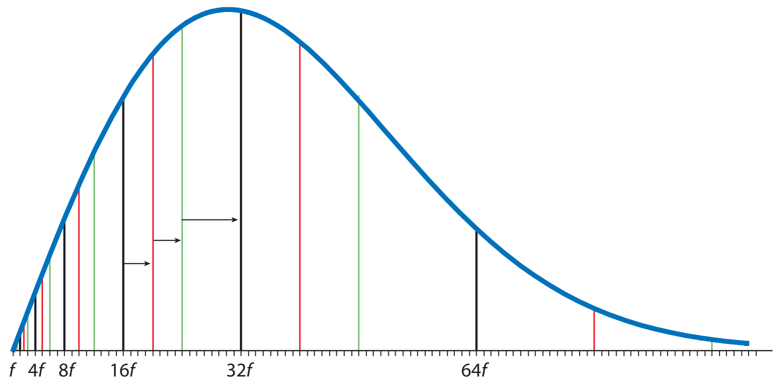
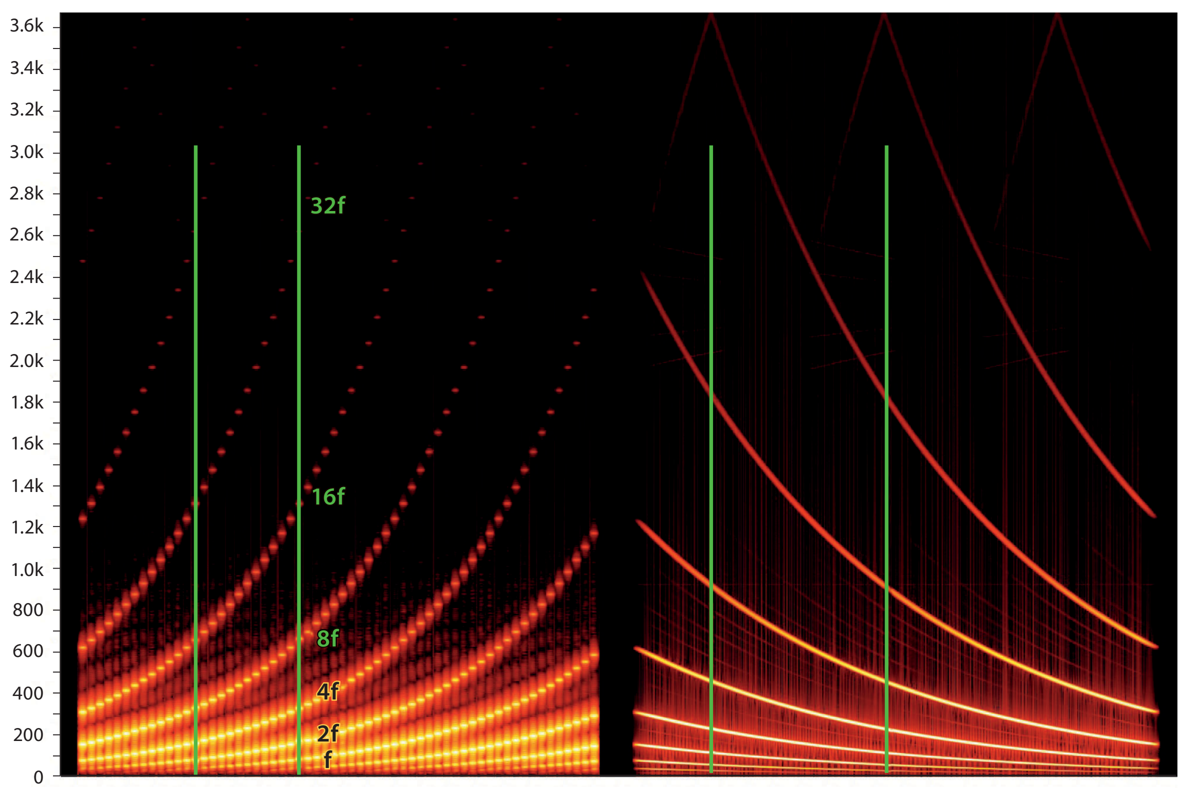


Figure 23.16

The Shepard tones scenario is shown here as successive upward progression of all the frequencies with amplitudes modulated by an envelope function. Only a few larger intervals are shown for clarity. After 12 semitones, the original black partials are exactly replaced; thus the tone has returned exactly to its former self. However, each of the semitone steps is an unmistakable upward change in pitch.

Figure 23.17

A sonogram revealing the structure and plan of the continuously and forever rising Shepard tone (left) and falling glissade Risset tone pitch illusions. Notice the self-similarity of the sonograms over time; these sound progressions could indeed continue forever without changing. The vertical lines reveal the fundamental repeated unit of the pattern.



that partial, according to the envelope, which modulates the amplitudes as $a_n = E(f_{m,n})$. Specifically, this envelope has the property $E(f_0) = 0$—that is, it is 0 at frequency f_0. This makes the $m = 0$ and $m = 12$ amplitudes *and* frequencies the same, so after an octave of semitone steps we have arrived back where we started, yet every semitone step is a rise in pitch. Figure 23.16 shows the spectrum and the envelope. We can also see the scheme in a sonogram (figure 23.17). After one octave rise in pitch, the pattern of amplitudes and frequencies is exactly where it started, which is confirmed by the sonogram. The vertical lines in figure 23.17 reveal the fundamental repeated unit of the pattern.

Shepard Tones and Autocorrelation

The autocorrelation reveals the secrets of the Shepard tones in another way. The scale and the steps are given in equation 23.4. We pick one of the tones in the sequence (figure 23.18).

Using the *Mathematica* CDF player you can download the *Shepard Tones* Wolfram Demonstrations Project applet from the wolfram.com website. This allows you to play the tones in any order you choose. The 12 buttons are arranged in a circular pattern, which is appropriate since the pattern exactly repeats at 12 intervals.

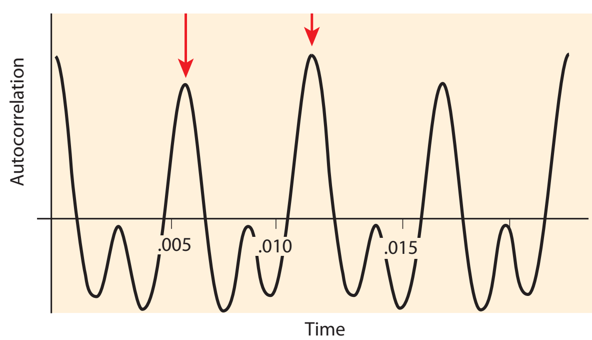


Figure 23.18
Autocorrelation of the sound for one stage of the Shepard series. The two arrows show autocorrelation peaks that could be expected to give pitch sensations. In the next (nominally higher) tone in the Shepard series, both of these peaks move to the left, which raises their pitch. This can continue indefinitely, since the innermost peak loses amplitude as it approaches short times, becoming insignificant. The next peak takes its place, and peaks farther out that were too distant at first to capture our attention move left to gradually take the place of the peak that originally controlled the pitch. Even with this scheme, some notes are qualitatively different from others: some give a stronger sensation of two pitches; others reveal, with practice, the low frequencies sneaking in to eventually take the place of higher ones when they approach $t = 0$.

Every step clockwise to the next note is an apparent semitone higher in pitch. There is a left shift of the autocorrelation peaks (which means they appear earlier in time and correspond to higher pitch), and there are small changes in their shape and height. The shift is just that required for the pitch to rise a semitone. When the first two tall peaks are about equal in height, they are an octave apart and both can be heard. Both tones rise in the next clockwise step up (both peaks shift left), but the peak closer to $t = 0$ (higher pitch) starts to diminish in height, gradually making the lower pitch more dominant even though each clockwise step raises the pitch of both peaks by moving them left. Eventually, new peaks moving from the right to left arrive, after 12 steps, to exactly reproduce the starting autocorrelation function.

Can the Shepard effect be achieved without such careful parsing of partials? The answer is yes, although perhaps not quite so convincingly. The basic idea is to use many notes across a wide frequency range, with the highest and lowest notes muted and the loudest notes in the middle of the range. Play successive rising (or falling) intervals while fading in or out notes at the extremes. Shepard-like effects have been used by the rock band Queen in the song "Tie Your Mother Down," have been exploited by Pink Floyd, and appear in the works of Bach and Chopin. Risset constructed a continuous glissade version of Shepard's discrete tones (figure 23.17, right).

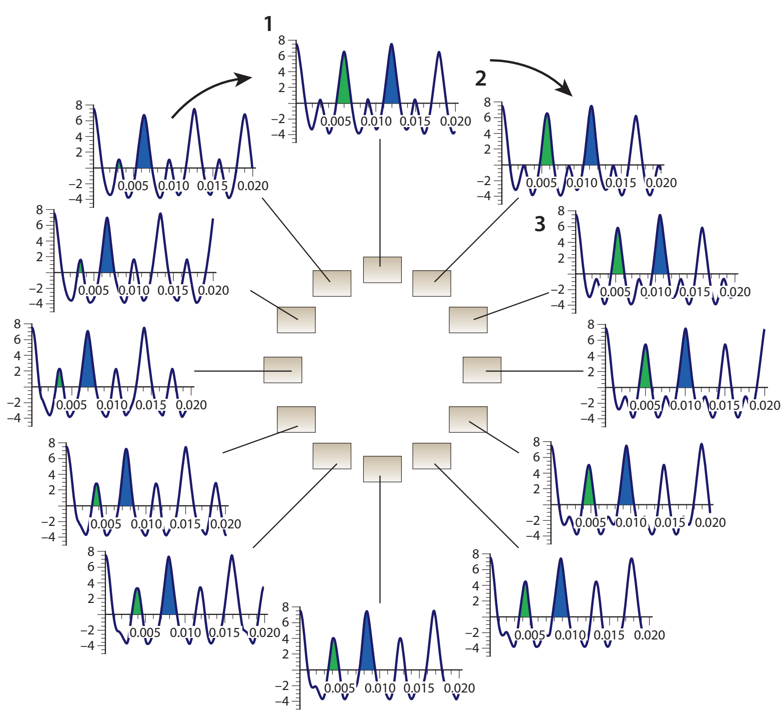


Figure 23.19

The Shepard tone illusion from the point of view of the autocorrelation functions, shown here for the notes of the *Shepard Tones Mathematica* applet. The scale is described in equation 23.4. Every clockwise step to the next note is a semitone higher in pitch. Every clockwise step (1 $\rightarrow$ 2, 2 $\rightarrow$ 3, and so on) gives a left shift of the autocorrelation peaks (which means that they appear earlier in time and correspond to higher pitch) and small changes in their shape and height. The shift is just that required for the pitch to rise a semitone. When the first two tall peaks are about equal in height, they are an octave apart and both can be heard. Both tones rise in pitch at the next clockwise step (both peaks shift left), but the peak closer to $t = 0$ (higher pitch) starts to diminish in height, gradually making the lower pitch more dominant even though each clockwise step raises the pitch of both peaks by moving them left. Eventually, new peaks moving from the right to left arrive, after 12 steps, to exactly reproduce the starting autocorrelation function.

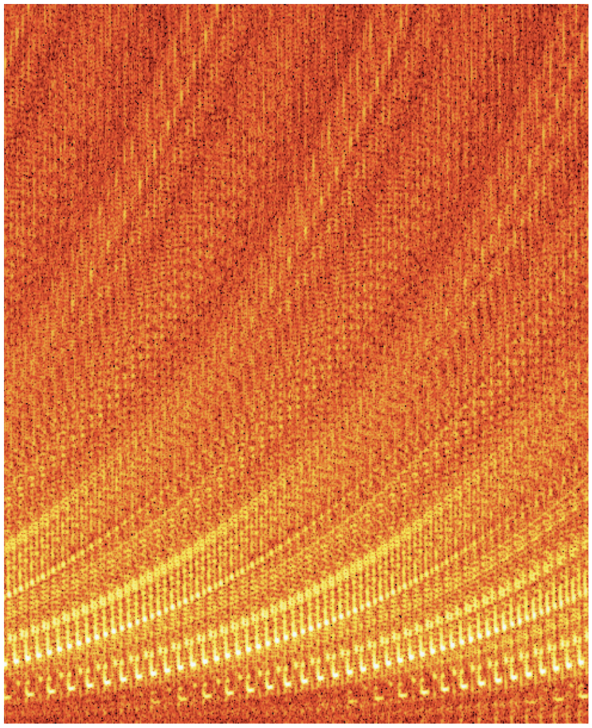


Figure 23.20
The same pattern of exactly repeated intervals of rising frequency seen in the Shepard tones are seen here in a sonogram of the Risset rhythm (RissetRhythm).

The evolution of the autocorrelation as the tone progresses through 12 steps "up" is shown in figure 23.19.

As Risset also realized, the general ideas behind Shepard tones can be applied to rhythm by using several percussionists, bringing in the slowest beating softly, everyone speeding the beat up from moment to moment, and fading percussionists out as their beat gets very rapid (figure 23.20). A percussionist who has been thus eliminated returns to soft slow beating, and so on. In section 23.22, we suggest that the concept of pitch be extended well below the nominal 20 Hz lower frequency limit of human hearing. In this light, the Risset beats and Shepard tone phenomena are the same—both are playing the same game with pitch.

A sonogram of the sound file (RissetRhythm.mp3, on whyyouhear-whatyouhear.com) reveals the self-similar rising pattern familiar from the Shepard tones (figure 23.20).

23.16 Chimes: Pitch without a Partial

We return to the perceived pitch, or *strike note*, of bells and chimes, which vibrate at many frequencies, just as a plucked string does. The clapper of a bell excites many modes at once; a complex vibration of the bell ensues that is a linear superposition of all these modes, each of which corresponds to a pure sinusoidal partial. If a mode has an antinode where the clapper hits, it tends to be strongly excited, and if it has a node there, it will be silent. This is the same principle as a plucked string, the difference being that the partials of the bell are not evenly spaced. There is no definite period of the resulting tone, and no unambiguous frequency that is the inverse of this period, yet "true" bells have a definite pitch—after all, they have to be able to ring out a tune.

The perceived pitch of a bell is usually not among the partials present. To achieve a pleasing tone, the partials cannot be placed helter-skelter. It is still an art to make a great-sounding church bell. The first, or lowest, partial tone is called the *hum tone* and is the simplest ellipsoidal vibration of the bell, in which the bell oscillates in the same way as the bowl shown in figure 20.10.

Many people find the unequally spaced partials in a bell or chime easier to hear out than the equally spaced partials in a periodic tone. We take as an example the sound file Strike Note of a Chime from the Acoustical Society of America's audio demonstration disk.[8] In this demonstration, the same chime tone is repeated nine times. The first time, we hear the chime; the next seven repeats are preceded by a pure sine tone at successive (inharmonic) partials contained in the chime tone. Although a trained ear can certainly hear the individual partials in the original tone with no help, the tendency on first hearing the chime is to listen holistically, taking in its pleasing timbre and hearing a definite overall pitch. However, after an individual partial is played, it is impossible for most people not to hear that particular partial ringing strongly in the subsequent chime tone, even though the chime playback is identical to the ones that preceded it. In effect, we are forced by the playing of the pure partial to hear the subsequent chime tone analytically. The last repetition of the chime is followed by the pure sine tone at the pitch of the chime. The pitch does not coincide with one of the chime's partials.

The Hosanna Bell in Freiburg

The Hosanna Bell in Freiburg, Germany, was commissioned in 1258. It is of a design now considered antiquated; its partials are not well spaced by modern criteria, owing ultimately to its shape. The bell is "long waisted" and "shaped like a large flowerpot with a heavy rim" according to William A. Hibbert, whose excellent 2008 PhD thesis (The Open University, Milton Keynes, United Kingdom) on bells includes a

[8] Houtsma, Rossing, and Wagenaars.

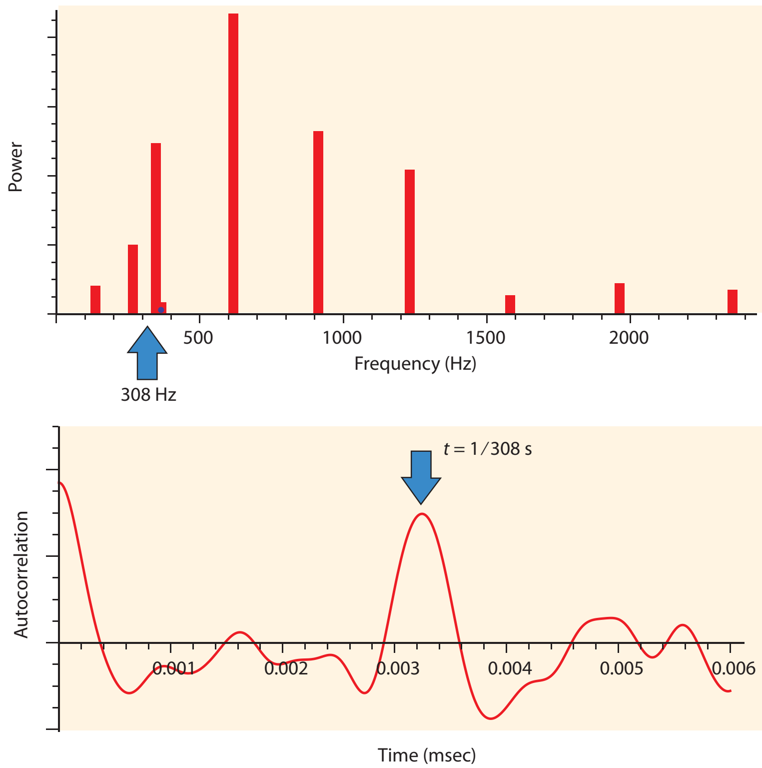


Figure 23.21

The power spectrum and autocorrelation for the Hosanna Bell in Freiburg, Germany, as rung normally by its clapper. There is a distinct peak in the autocorrelation at $t = 1/308$ s, corresponding to a 307.3 Hz average pitch reported by observers. The residue pitch formula 23.1 yields 308.6 Hz.

study of the Freiburg Hosanna. The Hosanna has a very definite pitch, which most people agree is near 308 Hz, rather high for a bell of this size. The partials are at frequencies 135.4, 267.4, 346.4, 365.8, 615.8, 912, 1231.6, 1582, 1962, 2356 Hz, with relative amplitudes 0.28, 0.5, 0.82, 0.1, 1.1, 0.85, 0.75, 0.2, 0.3, 0.25. Figure 23.21 shows the power spectrum and the autocorrelation using just this data. The residue pitch (formula 23.1) using $N_n = (0, 1, 1, 2, 2, 3, 4, 5, 6, 8)$—arrived at by counting the nearest whole number of periods of each partial present—is 308.6 Hz.

Pitch of a Kettle Drum

A well-struck kettle drum might have partials at 128, 192, and 256 Hz, which "should" give a residue pitch of 64 Hz, since these frequencies are all multiples of 64 Hz. However, almost everyone reports a pitch of 128 Hz instead. The 128 Hz component may dominate, but if the next two have a reasonable amplitude, a constant tone (as opposed to the kettle drumstrike) with these components does have a pitch of 64 Hz (although 128 can also be heard). The kettle drum, however, stubbornly seems to be 128 Hz.

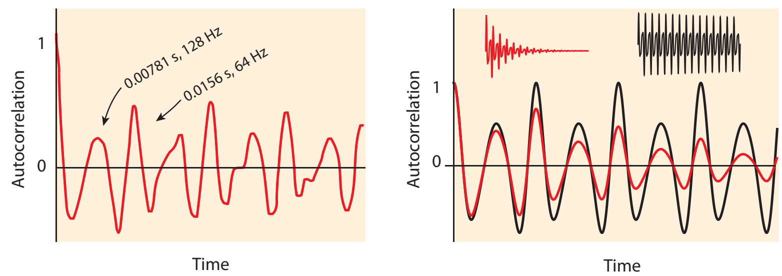


Figure 23.22
(Left) The autocorrelation function of the sound of a single strike of a kettle drum. Peaks corresponding to 64 Hz pitch and 128 Hz pitch are competing for dominance according to the marquee effect principle (see section 23.14). For most listeners, the pitch reported is 128 Hz. (Right) We synthesized a summary version of the tone, having only 128, 192, and 256 Hz components. Two sound traces and two corresponding autocorrelations are displayed. One tone was cut off rapidly, the other, less so. The pitch of the weakly damped tone when played over a speaker system is indeed often perceived an octave below the strongly damped one, as Rossing predicted, even though they differ only in how fast they are cut off. The autocorrelation gives some support to this impression: we note that for the longer lasting tone, the 64 Hz peak is taller and more prominent relative to the 128 Hz peak.

Figure 23.22 shows the autocorrelation for a recorded strike of the kettle drum. We see that the competition for “first billing” on the marquee is set up (see section 23.14) with an earlier peak at 128 Hz and a later but taller peak at 64 Hz. Apparently, here the earlier one wins.

Rossing speculated that the short duration of the kettle drum strike had something to do with the 128 Hz perception. To test this, we create artificial kettle drum strikes, and check whether the autocorrelation measure of pitch might lend support to this idea. Using amplitudes (1, 0.6, 0.3), in that order, for the 128, 192, and 256 Hz partials, we listen to the result for various exponential damping rates. A very interesting trend emerges: a short cutoff of the sound does cause the dominant pitch to rise an octave. Moreover, the autocorrelation measure confirms or at least makes plausible this trend, showing that the peak corresponding to a 64 Hz pitch becomes more prominent relative to the earlier 128 Hz peak as the tone is lengthened. (See figure 23.22, right, and listen to shortkettle.wav and longkettle.wav, available on whyyouhearwhatyouhear.com.)

23.17 Repetition Pitch

Noise is a common companion out-of-doors. The rustle of leaves, the sound of a waterfall, waves on a beach, feet shuffling along the ground are all noise sources. We now have many additional sources of outdoor noise, such as jet aircraft, cars passing by, and the general din of cities. The power spectrum of such noise is often not gathered into many sharp peaks, but rather diffused over very broad frequency ranges. If an average is taken over a long time, the power spectrum is a smooth continuum.

If a complex periodic tone of frequency f and period $T = 1/f$ is time delayed by half of its period—a time $T/2$—and added to its original self, the pitch of the tone will go up by an octave. The lowest partial, and in fact every odd (the 3rd, the 5th, and so on) partial above it, is nullified if it is added to itself half a period later: these partials are always the negative of themselves half a period of the fundamental earlier. For a 100 Hz tone, this is a delay of 0.005 s. If all the odd partials in a 100 Hz tone with 100, 200, 300, 400, . . . Hz partials are killed, leaving 200, 400, 600, . . . Hz, a 200 Hz pitch results (an octave higher). This fact figures in our explanation of Lord Rayliegh's *harmonic echo*, wherein he heard an echo of a woman's voice return at an octave higher than it left (see section 28.4). The autocorrelation of the signal reveals the repetitions as peaks and valleys shaping the power spectrum at the receiver accordingly (see figures 21.3 and 21.4). We discuss commonly encountered examples of this next.

Huygens at Chantilly

Very likely the first discovery and explanation of repetition pitch was provided in 1693 by Christian Huygens at the castle of Chantilly in France. Huygens is renowned for his theories of wave propagation; we encountered him in connection with refraction (see figure 2.9 and surrounding discussion). He noticed that the sound of a fountain located near a large set of steps is colored by a dominant pitch, and he correctly surmised that the reflections of the fountain noise by the nearby steps caused a repetitious echo consisting of a sequence of small echoes, one from each step. For 1/2 meter separating the steps, the echo from a hand clap near the fountain would send back echo pulses 340 to the second, giving a frequency of $f = 340$. The sound of the fountain is noisy, but the pitch can be heard nonetheless. The fountain noise can be thought of as thousands of little claps per second, each of which gets a repetition pitch echo. Huygens's own observations are remarkably modern, and his way of explaining the pitch that is heard is worth reading:

> When one is standing between the staircase and the fountain, one hears from the side of the staircase a resonance that possesses a certain musical pitch that continues, as long as the fountain spouts. One did not know where this tone originated from or improbable

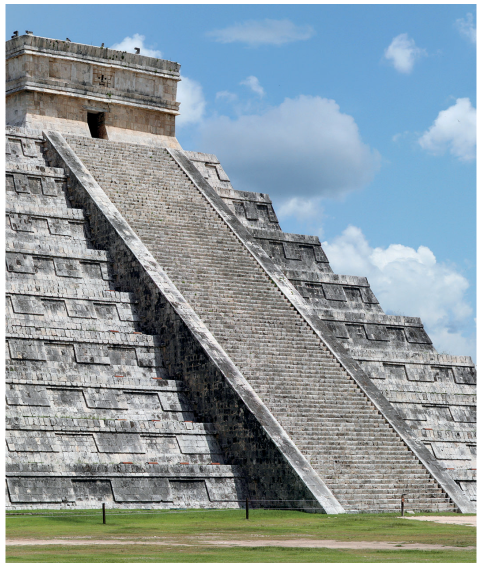


Figure 23.23
The Stairway to Heaven, Temple of Kukulkan, Chichén Itzá, Mexico. Courtesy Daniel Schwen, Creative Commons Attribution—Share Alike 3.0 Unported license.

explanations were given, which stimulated me to search for a better one. Soon I found that it originated from the reflection of the noise from the fountain against the steps of the staircase. Because like every sound, or rather noise, reiterated in equal small intervals produces a musical tone, and like the length of an organ pipe determines its own pitch by its length because the air pulsations arrive regularly within small time intervals used by the undulations to do the length of the pipe twice in case it is closed at the end, so I imagined that each, even the smallest, noise coming from the fountain, being reflected against the steps of the staircase, must arrive at the ear from each step as much later as the step is remote, and this by time differences just equal to those used by the undulations to travel to and fro the width of one step. Having measured that width equal to 17 inches, I made a roll of paper that had this length, and I found the same pitch that one heard at the foot of the staircase.[9]

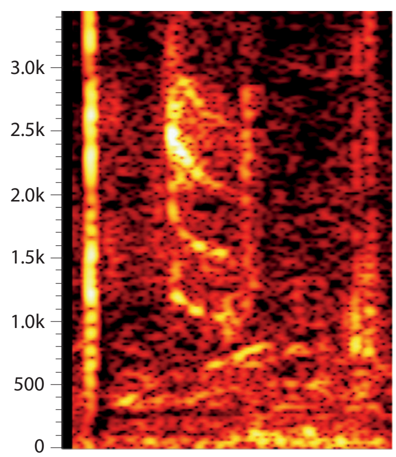


Figure 23.24
Sonogram of a handclap and return echo at the Stairway to Heaven, Temple of Kukulkan, Chichén Itzá, Mexico, on the Yucatan Peninsula, which comes back in the form of a chirp.

Temple of Kukulkan, Chichén Itzá

There is a famous chirp echo at the stairs of the Temple of Kukulkan, Chichén Itzá, Mexico (figure 23.23). The downward trending chirp can be seen in the sonogram in figure 23.24. A simulation in *Ripple* shows in detail how the chirped echo forms and how it differs depending on the

[9]Translated by Frans A. Bilsen, *Nederlands Akoestisch Genootschap*, 178 (2006), 1–8.

position of the listener. The running simulation is shown in figure 23.25, where individual reflections from the stairs are heading back to the source in the lower left corner.

The geometrical reason for the chirped echo is revealed in figure 23.26. The pulses returning to the source at the bottom of the stairs have diffracted off the top edge of each stair. The sound pulses arrive after a round-trip from source to edge and back. Therefore, the time delay between a given returning pulse and the next one from the stair above it is twice the difference in distance divided by the speed of sound. It is seen from the inset in figure 23.26 that the difference goes from being close to a, the step width, for sound coming from near the bottom of the stairs, to close to $(a^2 + h^2)^{1/2}$, where h is the rise of each stair, near the top of the stairs. This increased time delay for pulses arriving later accounts for the lower pitch.

The frequency of the chirp and its evolution with time clearly depend on the geometry of the stairs and the location of the source and listener. It is interesting to consider what a listener would hear standing on a platform 10 m above the source at the ground. It would be quite different in some respects!

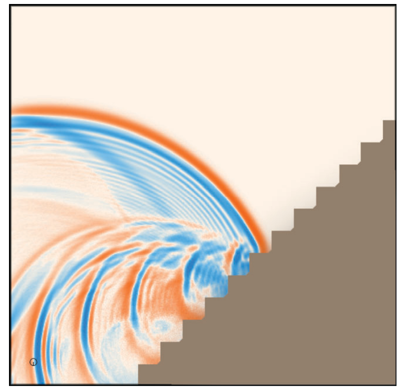


Figure 23.25

Ripple simulation of the chirped echo at the Stairway to Heaven, Temple of Kukulkan, Chichén Itzá, Mexico. The sound source was a sudden pulse at the lower left; the circular pressure wave emanating from the source can be seen traveling upward above the stairs, while successive reflections of that wave off the stairs are returning to the region of the source.

Ground Reflections

In fact, it takes just one repetition to give the sensation of pitch. You may have heard this many times without realizing it. Standing on hard ground or perhaps next to a building, sound can reach you by two different paths from a source. One is on a straight line through the air, and the other takes a single bounce before reaching your ears. Whatever you hear on the first path is quickly repeated on the second, with a delay controlled by the extra distance the sound travels on the second path and the speed of sound. Suppose the sound source is essentially white noise—for example, a jet overhead. White noise is characteristic of air jets and turbulence, and is an important source of sound, including in speech. (We discussed jet noise in section 14.7.) The perceived pitch f is given by the reciprocal of the delay time τ, $f = 1/\tau$. If there is a 0.01 second delay of the bounced sound, a 100 Hz pitch can be heard, which, however, is far from being a musical tone. In fact, one quickly becomes accustomed to the 100 Hz coloration, and it helps to change the sound by moving closer to or farther from the ground, or by moving the source itself, changing the delay and the pitch. One way to demonstrate the presence of repetition pitch is to play a little tune with it.

Figure 23.27 is a sonogram of a sound file that consists of a short segment of white noise, followed by the same white noise added to a copy with a 10 ms delay, and then next with a 20 ms delay, followed by a 40 ms delay, and last an 80 ms delay. The downward jumps in pitch can easily be heard.

The sonogram in figure 23.27 shows the antiresonance "notches" cut into the uniform white noise spectrum seen at the left of the figure. For the sound of a jet passing overhead (found in the sound file Jet Airplane Passing

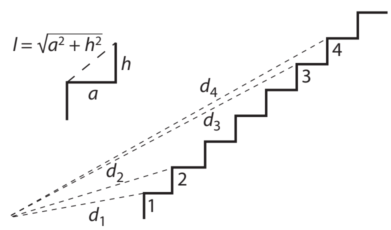


Figure 23.26

The geometry of the chirped echo at the Stairway to Heaven, Chichén Itzá, on the Yucatan Peninsula in Mexico. Successive rebounds from the edges of the stairs, at intervals depending on the distance increase from one stair to the next. Starting with a sound source at the left, the round-trip travel distance to the first stair edge 1 differs from that to the second stair edge 2 by an amount very slightly greater than $2a$, or twice the width of the stairs. The later part of the echo coming from farther up the stairs, however, has a path length difference of almost twice the hypotenuse $\ell = (a^2 + h^2)^{1/2}$, or about 40% greater.

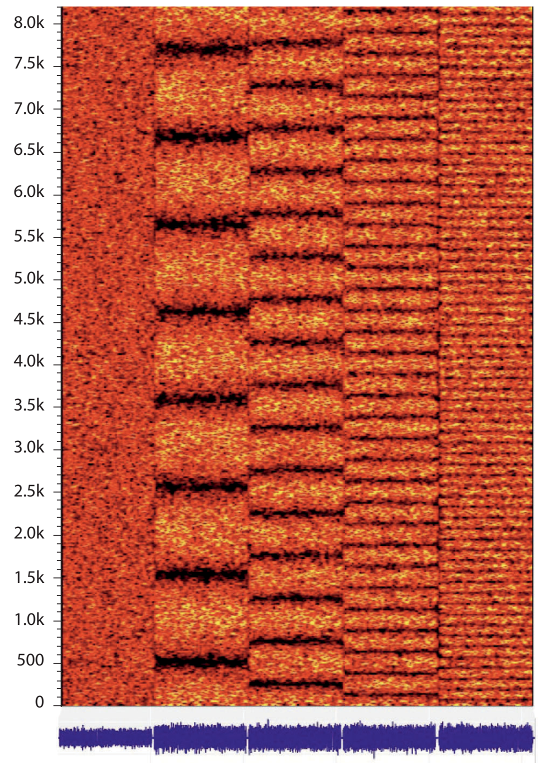


Figure 23.27

Sonogram of a white noise sample (left strip), followed by the white noise sample with a copy added with a 1 ms delay, then with a 2 ms delay, a 4 ms delay, and finally an 8 ms delay. The notches (and enhancements between) in the power spectrum caused by the repetition are clearly seen. The reader should verify that the frequencies of the notches are in the expected positions.

Overhead available at whyyouhearwhatyouhear.com), the sonogram of which is shown in figure 23.28, the notches are also clearly visible, but are now continuously changing with time. Starting at the left, as the jet approaches, the interval between the notches is decreasing and reaches a minimum when the jet is directly overhead. After the jet passes, the interval between the notches increases again. The effect is first a falling pitch, followed by a rising pitch after the jet passes. This effect is not like the Doppler effect: after the jet passes, the pitch would continue to fall if this were a Doppler phenomenon (see figure 7.31 or 7.32). If you look carefully, the whine of the jet engine is a faint trace in the upper part of the sonogram that does have this Doppler signature. You can hear the whine of the engine and the Doppler effect separately from the repetition pitch.

In order to hear these repetition pitch effects, the listener must be standing on the ground or at least on some hard surface of considerable extent, like the top of a parking structure. You would not hear the effect from an apartment balcony.

We can estimate the repetition pitch as the jet passes overhead by computing the path length difference between the direct and bounce paths (see figure 23.29). To get both lengths, we need only apply the Pythagorean

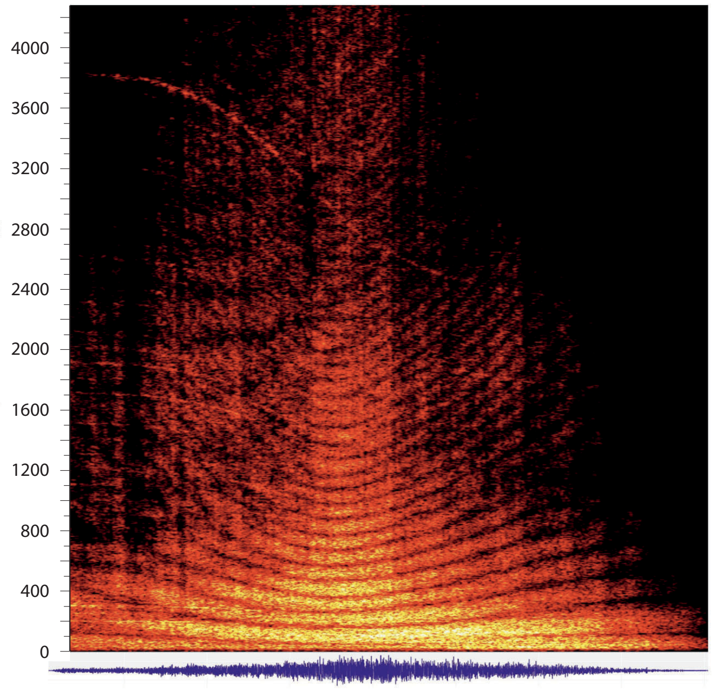


Figure 23.28

Sonogram of a jet passing overhead. The recording was made with a microphone placed about 1.8 m above the ground. The repetition pitch falls to 96 Hz when the jet is overhead; it is higher both before and after. The geometry is such that the time delay for the bounce sound is at a maximum with the jet overhead. Thus, the frequency of the perceived repetition pitch is at a minimum. The Doppler effect is also at work here, seen in the frequency change of a high-pitched whine of the engine turbine, which starts above 3600 Hz as the jet approaches. It declines to around 2400 Hz as the jet recedes; this pitch is falling over the entire time interval, and it falls the fastest as the jet is overhead. This is in contrast to the repetition pitch, which rises after the jet passes.

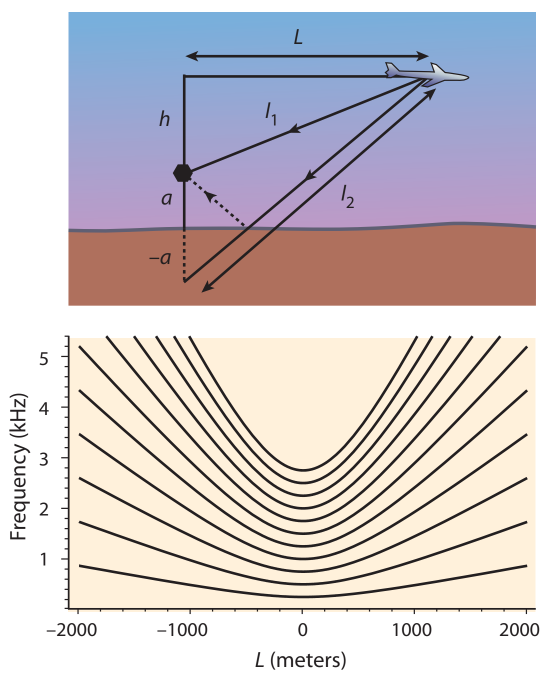


Figure 23.29

Geometry of sound reaching a microphone a distance *a* above ground by two paths, of length ℓ_1 and ℓ_2. The path ℓ_2 is longer and, by bouncing once off the ground, gives rise to a time-delayed repetition of the sound. The length of the path ℓ_2 can be determined by extending the path along a line underground reaching a distance *a* below ground. The determination of path length ℓ_2 then involves a right triangle and the Pythagorean theorem, as does the path ℓ_1. The time delay is given by the difference between the two path lengths divided by the speed of sound. The repetition pitch that is heard is the inverse of this time delay.

theorem, since both lengths are the hypotenuse of right triangles. We have $\ell_1 = \sqrt{L^2 + h^2}$, $\ell_2 = \sqrt{L^2 + (h + 2a)^2}$. The lines plotted in figure 23.29 are then

$$f_n(L) = \frac{nc}{(\ell_2 - \ell_1)}; \quad n = 1, 2, \ldots, 11, \tag{23.5}$$

where c is the speed of sound. (A single repetition with a time delay T puts notches in the power spectrum at frequencies $f_n = (n - 1/2)/T$, as shown also in figures 21.3 and 23.27.)

Repetition pitch is far more commonly heard than we are normally aware of. For example, if you are in a room with a ceiling fan that makes a lot of broadband noise, and there is a hard floor that is very reflective of sound, try putting your head at different distances above the floor. You may hear a changing pitch. The next time a jet flies overhead, try positioning your head at different heights above the ground. You'll have control of the repetition pitch, which will go up as your head approaches the pavement.

Visually impaired people often use repetition pitch to judge distance. If you create a sound and there is a wall many meters away, you will hear a distinct echo separated in time, not a pitch. But as you come closer to the surface, reflected sound arrives very quickly, say, within a few milliseconds. The repetition pitch rises as the wall is approached, since there is a smaller time delay, much too quick to detect a separate echo. If you become sensitive to repetition pitch, it can be used to judge distance quite accurately.

23.18 Quantifying Frequency

Frequency can be measured in Hz, as accurately and with as many decimal places as you please, so you might imagine that there's no need for any other system. However, the way we actually hear intervenes and makes another way of measuring frequency much more useful. We are far more sensitive to small changes of a few Hz at lower frequencies than we are at higher frequencies. A 5 Hz shift in a 50 Hz tone is a 10% effect, but a 5 Hz shift in a 5000 Hz tone is a 0.1% effect. Most people can tell the difference easily between 50 and 55 Hz, but cannot tell the difference between 5000 and 5005 Hz. The best measure of frequency differences would reflect our sensitivity to them.

Cents

Any octave interval is divided into 1200 parts, called *cents*. An A3 at 220 Hz and an A4 an octave higher at 440 Hz, an absolute difference of 220 Hz, differ by 1200 cents. The division is, however, not into 1200 intervals of equal frequency difference, but rather equal intervals in the logarithm of the frequency. In the Western, 12-tone, equal-tempered system, this means there are 100 cents between adjacent keys on the piano keyboard. (We shall discuss various approaches to temperament in chapter 26.) We can express the measure of cents as

$$n(\text{cents}) = 1200 \log_2 \left(\frac{f_2}{f_1}\right) \approx 3896 \log_{10} \left(\frac{f_2}{f_1}\right)$$
$$= 3896 \log_{10}(f_2) - 3896 \log_{10}(f_1), \tag{23.6}$$

where n is the number of cents up in going from the lower frequency f_1 to the higher frequency f_2. Since $\log_2(2) = 1$, any two frequencies that differ by a factor of two are separated by 1200 cents.

Just Noticeable Difference (JND)

The *just noticeable difference*, or JND, is defined as the minimum pure tone frequency change required for a listener to detect a change in pitch. Humans are most sensitive to pitch changes in the range 50 to 2000 Hz; in this range, most of us can detect a change from 2 to 10 Hz.[10]

The JND is about 1/30th of the critical bandwidth throughout the range 20 to 10,000 Hz. Thus it is about 3 Hz for a 200 Hz sine tone, where the critical bandwidth is about 90 Hz, and it is about 70 Hz for a 10,000 Hz sine tone, where the critical bandwidth is about 2000 Hz. This suggests that the critical bandwidth is somehow involved in pitch detection. It also is suggestive of a *peak locating* algorithm. Suppose a 100 Hz signal is smeared out symmetrically, 25 Hz on either side of 100 Hz. Even with the smearing, the maximum is still at exactly 100 Hz; if some way exists to find the maximum response in the curve, then the resolution of frequency will be much better than ± 25 Hz. This trick is used to advantage in many fields where the source of data is known to be sharp, but the instruments smear it out. The orbital parameters of planets around distant stars are measured this way, for example. An active neural-hair cell feedback loop is known to sharpen pitch detection. See section 21.4.

Time or Place?

The controversies and the issues surrounding pitch perception and related phenomena are intimately connected to two seemingly disparate views of

[10]To determine your own JND, take the test at http://webphysics.davidson.edu/faculty/dmb/soundRM/jnd/jnd.html.

our hearing mechanism: One view puts primary emphasis on place and partials, meaning the association of location on our basilar membrane with corresponding sinusoidal frequencies. This view is centered on the idea of Fourier frequency analysis of sound. The other camp puts primary emphasis on the ability of our ear and brain system to sort out temporal information. Clearly, we can decode transient sound, or else we could not understand speech.

There are many problems with Ohm and Helmholtz's tendency to overrate frequency analysis in the context of human hearing. For starters, consider the beating of two partials close in frequency. If we really detect separate partials, why should we hear, as we do, the loudness beating of at a single average frequency rather than the presence of two partials? Transient sounds are natural to explain in the context of temporal theories of hearing but very difficult to explain within a place theory context based on partials. A sentence could be Fourier analyzed in principle; a broad band of power over a large range frequencies would result. (Try this on your laptop.) But this is hardly the way we hear.

Most sounds do not arrive in long-lasting tones, ripe for Fourier analysis. We clearly need to have a means of temporal processing. The temporal school of hearing asserts that we can process separate phenomena that occur as fast as 5000 times per second. The temporal theories are quite compatible with the notion, promoted here, that our sense of pitch arises from periodicity and autocorrelation of sound. This puts pitch perception up a notch in the auditory system, more neural than mechanical. We are not waiting for frequency detectors to send the first signals to the brain, but rather pitch and timbre are deduced from the temporal data.

The two schools of thought, time versus place, are not mutually exclusive. There is evidence that temporal theory applies to most of what happens below 5000 Hz, and place theory above 5000 Hz. Our neural processes cannot keep track of time intervals shorter than about 0.0002 second, or 5000 Hz, so it is reasonable that we lose the precision of timing above that frequency and switch over to a place mechanism for detecting frequency—the region of maximum excitation of the basilar membrane—above 5000 Hz. People with good hearing perceive frequencies of 20 to 15,000 or 20,000 Hz. However, frequency estimation is poor and the sense of pitch is next to nonexistent above 5000 Hz. There is no such thing as timbre above 5000 Hz either. It is doubtful that you can recognize tunes composed of frequencies entirely above 5000 Hz; this is true even of people with perfect pitch.

It is no coincidence that the grand piano's highest pitch is 4186 Hz (C8). No familiar musical instrument is played above about 4400 Hz. Complex tones consisting of many frequencies, all above 5000 Hz, are heard analytically as poorly resolved individual frequencies, rather than holistically as a tone with pitch.

23.19 Pitch Class, the Octave Ambiguity, and Perfect Pitch

There seems to be no agreement on how many people have perfect pitch, whether or how it can be learned or acquired, or most significantly even what it is exactly. At the very least, someone with perfect pitch can tell you the note being played on a piano with their eyes closed. They may be able to tell you that the A440 is flat by a quarter of a semitone. There are studies showing that people who speak Mandarin Chinese from birth are much more likely to have perfect pitch, the presumed reason being that the same sounds differing only in pitch have different meanings and training matters.

Having perfect pitch is like color perception is to most of us: perfectly natural, done without a thought. It does not seem to be a great accomplishment to match the color of a banana on the color circle. People with perfect pitch report a G sharp just as matter-of-factly. Indeed, perfect pitch has its downside. Most people don't notice if a tune is transposed up a full tone. This might be as unsavory as a blue banana to someone with perfect pitch.

Perfect pitch does not extend to the point that its possessors can tell you the pitch is 407 Hz, and not 408 or 406 Hz. Perhaps more surprisingly, perfect pitch applies only to pitch classes—that is, the note may be a G3, but G4 is reported. It makes sense to place color hues in a circle, starting with red and moving through orange, yellow, green, cyan, blue, magenta, and back to red, completing the circle with no jumps or discontinuities in hue. Perfect pitch then corresponds to naming hue, which is to say, the angle on the color circle. We can imagine colors laid out in a spiral, with each successive octave lying above the other less color saturated than the last. All the hues cycle in the same way each octave. Asking someone to name both the hue *and* the saturation "octave" is clearly more difficult, and involves the eye's color receptors and the brain in a different way. The *pitch spiral* (figure 23.30) illustrates the similarity of members of the same pitch class. The pitch spiral should be applied only to pure sinusoidal tones; otherwise, many other ambiguities may arise.

Complex tones can suffer much more pitch ambiguity than simple ones. We have seen how sensitive pitch is to slight systematic deviations in the Seebeck experiment (section 23.14), dropping the perceived pitch an octave with tiny shifts in the tone. Because of the partials they share in common, a complex tone C2, and a tone an octave above it, C3, have a similarity that is quite striking. In fact, certain complex tones could be parsed into both a C2 and a C3 note, but this might be up to the listener; there would be no "correct" answer. All notes separated by octaves are defined to be in the same *pitch class*. It is clear, however, that just as with partials in a complex tone, which can either blend into the whole or be heard out, voluntarily or not, attention can be called to the

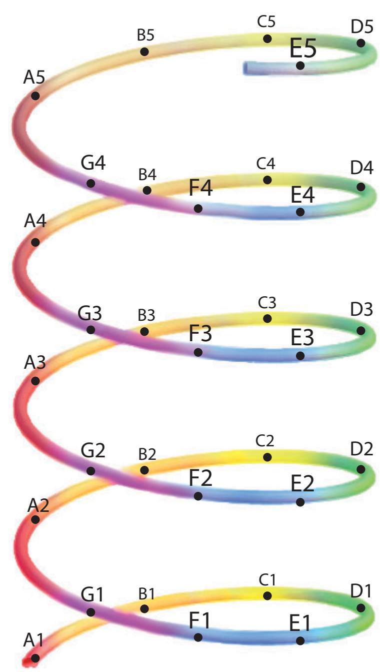


Figure 23.30

Pitch spiral, showing the similarity (here schematically seen as color and proximity in space) of members of the same pitch class, which are "different" in one way but the same in another.

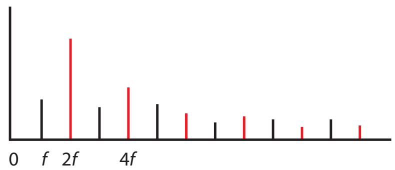


Figure 23.31

Is this one tone with a pitch *f*, or is it two, the black with pitch *f* and the red with pitch 2*f*?

possible existence of two notes by musical context, learning, and other cues.

In figure 23.31, we see the spectrum of an ordinary complex periodic tone, with every other partial colored red, beginning with the second partial. This tone "ought" to have an unambiguous, strong pitch at frequency f. Yet who is to say that the tone is not a combination of two tones? The lower tone consists of the black power spectrum, and the upper tone the red power spectrum. Both tones are complex, one an octave above the other. A tone like this could be produced with a combination of a clarinet (lacking its even partials) playing a G3 (black spectrum) and an oboe an octave above, playing a G4 (red spectrum), with all its overtone partials present. A clever experiment was described in a Stanford PhD thesis by S. McAdams.[11] The harmonics of an oboe were separated into even and odd components, and then an independent vibrato was added to just the odd components. This results in the sensation of two independent notes on two different instruments, separated in pitch by an octave.

23.20 Parsing and Persistence: Analytic versus Synthetic Hearing

We have mentioned switching between analytic and synthetic listening and how this can be influenced by context. An example was the chime sound, analyzed into partials, as heard in the file Strike Note of a Chime from the Acoustical Society of America's audio demonstration disk or as heard when a partial in a complex tone in *Fourier* or MAX *Partials* is adjusted, calling attention to it. Two intriguing questions are raised: Does auditory persistence apply to two or more partials at once? Does this work in a musical context? These questions were addressed by Helmholtz 150 years ago:

> Get a powerful bass voice to sing E flat to the vowel O, in *sore* (more like *aw* in *saw* than *o* in *so*), gently touch B flat on the piano, which is the 12th or third partial tone of E flat, and let its sound die away while you are listening to it attentively. The note B-flat on the piano will appear really not to die away, but to keep on sounding, even when its string is damped by removing the finger from the digital, because the ear unconsciously passes from the tone of the piano to the partial tone of the same pitch produced by the singer, and takes the latter for continuation of the former. But when the finger is removed from the key, and the damper has fallen, it is of course impossible that the tone of the string should have continued sounding.[12]

[11] Stanford University, Dept. of Music technical report, STAN-M-22 (1984).

[12] *On the Sensation of Tone*. There is a spectacular modern reversal of this experiment, in which a computer controlled piano with dampers down is able to reproduce recognizable speech. Look up "Speaking Piano" on YouTube.

Memory of the partials that were very recently present does seem to color our perception of a tone. This is why the strike note of a chime or bell is so important. Some partials die off faster than others, and memory of them may color what we hear two seconds into a chime tone, making partials that are very weak have a disproportionate effect.

The boundaries between analytic and synthetic listening are not rigid. Are we listening analytically or synthetically when we parse a compound sound into several distinct musical instruments and perhaps many distinct simultaneous notes? Or, when we recognize two singers in unison as such, rather than a single voice? (Depending on the voices, the context, vibrato, and so on, this may be quite difficult to do.) It is one form of analytic listening to hear out individual partials, but perhaps a not so distantly related form to hear out patterns of several partials that might indicate the presence of an oboe, for example, among other instruments playing simultaneously. According to this definition, normal listening is part analytical and part synthetic, even when individual partials are not heard.

23.21 Deutsch's Octave Illusion

An auditory illusion was discovered by Diana Deutsch in the 1970s and is rich with implications. It is perhaps the most remarkable audio illusion yet discovered. (Unless you happen to be one of the few who hears what is happening correctly.) One of the amazing aspects of this illusion is that different people hear it in starkly different and easily describable ways. There are no nagging ambiguities, no mincing of interpretations or words, which seems to happen too often in pitch perception. This illusion might be the key to one of the difficulties in resolving pitch in phantom tone controversies: What if people aren't hearing even approximately the same thing?

We quote from Deutsch's website:

> Two tones that are spaced an octave apart are alternated repeatedly at a rate of four per second. The identical sequence is played over headphones to both ears simultaneously, except that when the right ear receives the high tone the left ear receives the low tone, and vice versa. The tones are sine waves of constant amplitude, and follow each other without amplitude drops at the transitions. So in fact the listener is presented with a single, continuous two-tone chord, with the ear of input for each component switching repeatedly.
>
> Despite its simplicity, this pattern is almost never heard correctly, and instead produces a number of illusions. Many people hear a single tone which switches from ear to ear, while its pitch simultaneously

> shifts back and forth between high and low. So it seems as though one ear is receiving the pattern 'high tone—silence—high tone—silence' while at the same time the other ear is receiving the pattern 'silence—low tone—silence—low tone.' Even more strangely, when the earphone positions are reversed many people hear the same thing: The tone that had appeared in the right ear still appears in the right ear, and the tone that had appeared in the left ear still appears in the left ear.[13]

Now, that's an illusion!

Pitch and Loudness

There are slight but measurable pitch changes with loudness. Pure tones of low frequency tend to go down in perceived frequency with increasing loudness, whereas tones of high frequency tend to rise. The downward shift maximizes around 150 Hz to between 0 and 75 cents when a 250 Hz pure tone is increased from 40 to 90 dB. The largest upward shift happens at about 8000 Hz. There is little or no shift for middle frequencies (1000 to 2000 Hz).

For complex tones, the amount of shift and its direction depends on which partials dominate the tone, but fortunately for music, complex tones shift much less than pure tones, as if the extra partials help to "anchor" the pitch. The reason may be contained in the previous paragraph: if low- and high-frequency partials are reacting in opposite directions, then a complex tone containing both might not shift at all, on average. It would be interesting to study partials that are heard out analytically in loud versus soft sounds: do partials in a complex tone shift as they would if heard in isolation?

23.22 An Extended Definition of Pitch

A wooden yardstick pressed firmly on a tabletop with a few centimeters protruding off the edge vibrates at say 60 Hz and so has a pitch of 60 Hz when "plucked." Let more of the stick protrude, and the vibration frequency slows to 10 Hz. The large number of harmonics of 10 Hz make the repetitious tapping sound quite audible in spite of the 0.1 s period. We can almost count the number of periods per second; at 5 Hz, we *can* count them. The 60 Hz frequency has become 10 Hz or 5 Hz. The sound is still

[13] See deutsch.ucsd.edu/psychology/pages.php?i=101.

audible as we slide from 60 to 5 Hz. At what point does it stop having a pitch?

The point is, it doesn't stop having a pitch. The pitch is 5 Hz. Once digested, this simple example shows that the residue pitch effect is a necessity, not an illusion. Our hearing must grade from tone into counting continuously with no abrupt changes. The residue pitch becomes the counting frequency, which is also necessarily the "pitch."

According to this definition, humans can hear the *pitch* of sounds for 0 Hz to about 5000 Hz, above which we lose a precise sense of pitch. This point of view highlights the difference between pitch and partials, which cannot be heard below about 20 Hz but can be heard (although not precisely in terms of estimating the frequency) above 5000 Hz.

This definition of *pitch* overlaps and extends the traditional use of the word in a musical context. It is consistent with the use of *pitch* in other contexts: the pitch of a screw, or the pitch of seating in an airliner. Pitch is always about the number of objects in a given space, or in music, time.

We illustrate this principle with the sound file 10HzMissingFund.wav (available at whyyouhearwhatyouhear.com), which is a summation of 25 harmonic (equally spaced) partials starting at 20 Hz—20, 30, 40, . . . Hz. Even better, set up MAX *Partials* with the fundamental and a few more lowest partials absent. Slide the base frequency to 10 Hz, and slide the mouse pointer over the higher partials to create a lump, or formant, of amplitudes. You should hear 10 Hz pulsing. Next, slowly raise the base frequency. Separate pulses that sound rather like helicopter blades idling at 10 Hz will become a continuous beat tone, with a pitch equal to the pitch of the residue. This process may not be complete until 100 Hz or above. The transition is smooth, however; nothing abrupt happens in the range 10 Hz to 100 Hz. Our hearing system grades continuously from the "counting" regime at 10 Hz and below to the regime of a tone well above 50 Hz.

24 Timbre

Timbre is the third of three executive summaries of sound provided to our consciousness, after pitch and loudness. Timbre is what is different about a trumpet and a clarinet when they (separately) play a 220 Hz tone at the same loudness. Like all sensory functions, timbre is a complex, difficult to measure, psychophysical phenomenon. As with the sense of taste, there are subtleties and there are also connoisseurs. Impressions of timbre depend on context and experience. We shall try to sidestep these nuances and take a simplistic view: timbre is a kind of executive summary of the distribution of amplitudes of the various partials in a complex tone. Two sounds having all the same frequencies present but differing in amplitude have a different timbre. It is easy to experiment with timbre differences by changing the power spectrum amplitudes in *Fourier* or MAX *Partials*.

It is not possible to characterize timbre on a single scale from low to high, as we can for pitch and loudness. Timbre depends on the relative amplitude of the various partials, but it is often hard to describe. Some universals do exist: tones consisting only of the odd harmonics sound "hollow."[1] Tones with many high upper partials may sound raspy or harsh, as in buzzing into a trumpet mouthpiece; those with few may sound mellow, as in a flute at low sound volume. Pure sinusoidal partials are dull and colorless, especially at low frequency.

24.1 Timbre and Phase

Shape Depends on Phase

When two sinusoids of frequency f_1 and f_2 are added together, the shape of the resulting curve depends on the relative phase ϕ:

$$y(t) = A_1 \sin(2\pi f_1 t) + A_2 \sin(2\pi f_2 t + \phi). \tag{24.1}$$

[1]For good reason: half-open pipes (that is, a hollow space) have only odd harmonics as resonances!

Figure 24.1
Both of the black traces are the result of adding the colored sinusoids together. These are the same in the two cases, except for a phase shift. Both black traces would sound almost exactly the same, and both have the same power spectrum.

The phase controls the offset of the crests and troughs of different partials, and affects the resulting shape when they are added together (figure 24.1).

Ohm-Helmholtz Phase Law

We can phase-shift a sine wave by adding a phase ϕ inside the argument of the sine—that is, $y(t) = A\sin(2\pi f t + \phi)$. We allegedly cannot hear one ϕ from another. In a complex tone, there may be many partials and many possible phases ϕ_i, which for a harmonic tone reads

$$y(t) = A_1 \sin(2\pi f t + \phi_1) + A_2 \sin(4\pi f t + \phi_2) + A_3 \sin(6\pi f t + \phi_3) + \cdots . \tag{24.2}$$

Now the question arises: can we hear differences as the ϕ_i are changed? There is a generalization of this question to any steady inharmonic tone like a chime. Here, we will consider mainly pairs of sinusoids that fall into both harmonic and inharmonic classes.

The autocorrelation (and thus the power spectrum) of a steady tone is unchanged by the phase of its partials, so it is no surprise that pitch is not affected by the phases. However, since the phases drastically affect the *shape* of the waveform, one might wonder if the timbre is affected. Although Ohm knew that the timbre of a note is a result of the amplitudes A_i, he assigned no importance to the phases ϕ_i. Later, Helmholtz made this an explicit principle, asserting that phases did not affect our perception of the tone. We shall call this the *Ohm-Helmholtz phase law*, which is a corollary of Ohm's law, since phases are not mentioned in the statement of Ohm's law.

Figure 24.2 shows a case with many partials; when the phases are randomized the waveform changes drastically. According to the Ohm-Helmholtz phase law, we can't hear the difference.

Are we in fact totally insensitive to the phases? Elaborate and ingenious experiments were devised to check this, in particular by Helmholtz and Rudolph Koenig. These experiments were not taken to be conclusive, because it could not be verified that other things besides phase had not changed. Now, we can check this more easily with computers, although some of the same pitfalls still exist. Primarily, it is important to ensure that there are no partials shared by tones that are to be phase-shifted. If in fact some are shared, then phase shifting will alter the *amplitude* of the shared partial, since the relative phase will induce constructive or destructive interference.

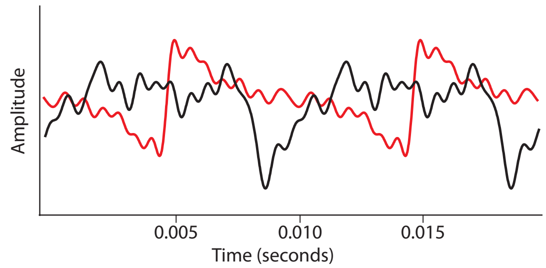


Figure 24.2

The two 100 Hz traces shown here have the same power spectrum. They differ in the phase choices, ϕ_k, in equation 24.2 in the 15 partials that are present. When played, they sound the same or nearly the same, supporting the Ohm-Helmholtz phase law. However, if you play them back at 1/4 speed (25 Hz), they sound significantly different.

Rationale for Insensitivity to Relative Phase of Harmonic Partials

There is more than one rationale for our phase insensitivity to harmonic partials in a tone. Even in musical sounds, relative phases of the partials can vary naturally over time. Real strings are an example, where the higher harmonics are not quite exact multiples of the fundamental—for example, the second partial of a string might be 401 Hz if the first partial is 200 Hz. The fact that the second partial is one Hz away from being perfectly harmonic means that the relative phase of the second partial is going full circle through 2π once per second:

$$\sin(2\pi\, 200t) + \sin(2\pi\, 401t) = \sin(2\pi\, 200t) + \sin(2\pi\, 400t + \phi(t)), \quad (24.3)$$

where the phase $\phi(t) = 2\pi t$. If we were quite sensitive to phase, the piano would sound very weird indeed. Certainly, however, the timbre of a piano is not perfectly constant after a note is struck with the pedal down. Some of its shimmering qualities may be due to slight changes of timbre due to phase drifts, although interaction of string and soundboard is also important.

The second rationale for our insensitivity to phase is that we need to hear more or less the same thing when in similar locations, as in two different seats in a concert hall. Two listeners sitting in different places relative to the sound sources can experience a marked shift in the relative phase of the partials they are hearing. However, they also experience different strength of the partials, due to reflections from walls and the ceiling that enhance certain frequencies at the expense of others, which we saw in connection with the repetition pitch effect (see section 23.17 and figure 27.17).

To simplify things, we go outdoors to eliminate most reflections. Suppose one singer is producing a pure partial at 100 Hz, and another several meters away is generating a pure partial a fifth above, at 150 Hz. Two listeners at different locations receive the partials with altered phases since the time delays are different. (The phase shift due to a difference d is distance is $2\pi d/\lambda$, where λ is the wavelength of the partial in question.) Figure 24.3 shows a numerical simulation of this situation, done in *Ripple*, resulting in a very different-looking sound trace for the two listeners.

Nonetheless, we expect them to hear essentially the same tone and timbre. We come to the conclusion that acute sensitivity to the relative phase of partials would not be a good thing. We would not want the two traces in figure 24.2 or 24.3 to sound as different as they look.

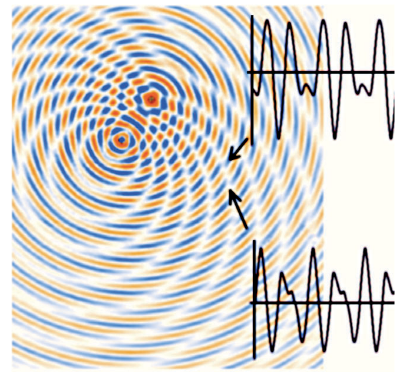


Figure 24.3
The sound waves emanating from two "singers," each singing a different purely sinusoidal tone (one at 100 Hz, and another at 150 Hz), are shown reaching two "listeners" some distance away. Because the listeners have different distances to the two singers, they experience a different relative phase of the 100 Hz and 150 Hz sine waves. The listeners receive different sound traces, but with very nearly the same power spectrum. The Ohm-Helmholtz law says that the listeners will hear essentially the same timbre.

24.2 Amplitude and Timbre Beats

The simplest form of beats are heard as undulations of amplitude (loudness) occurring between two sinusoidal partials if they differ by less than about 10 Hz. As the difference increases, the sensation of beats eventually gives way to a "fused" tone, which sounds "rough" when the frequency difference goes above 15 Hz. If we increase the frequency difference further, the fused tone remains rough but becomes discernible as two separate tones. Still further, the rough sensation finally becomes smooth. This happens when the two tones "get out of each other's way" on the basilar membrane: each pure tone affects a nonoverlapping *critical band*. (See section 26.1 for more details about these effects, which figure strongly in our sense of dissonance and therefore our choices of musical temperament.)

Two sinusoids combined, $f(t) = \sin(2\pi f_1 t) + \sin(2\pi f_2 t)$, can also be written as

$$
\begin{aligned}
y(t) &= \sin(2\pi f_1 t) + \sin(2\pi f_2 t) = 2\cos[2\pi(f_1 - f_2)t/2]\sin[2\pi(f_1 + f_2)t/2] \\
&= 2\cos[2\pi\Delta f t/2]\sin[2\pi\bar{f}t],
\end{aligned}
\tag{24.4}
$$

using the identity $\sin v + \sin u = 2\cos\frac{1}{2}(v-u)\sin\frac{1}{2}(v+u)$. *The two sinusoids may be written as a product of two new sinusoids, one with the average frequency* $\bar{f} = (f_1 + f_2)/2$ *and one with half the difference frequency* $\Delta f/2 = (f_1 - f_2)/2$.

The reader is encouraged to add two sinusoids differing by a few Hz to hear the beats. They are also quite visible on a plot (see figure 24.4). All this seems straightforward, but, in fact, subtleties abound when two sinusoids are added, especially when it comes to what we hear and what we don't hear.

The case of adding 27 Hz and 25 Hz sinusoids is shown in figure 24.4. The GCD of these two frequencies is 1 Hz, and this is indeed the true

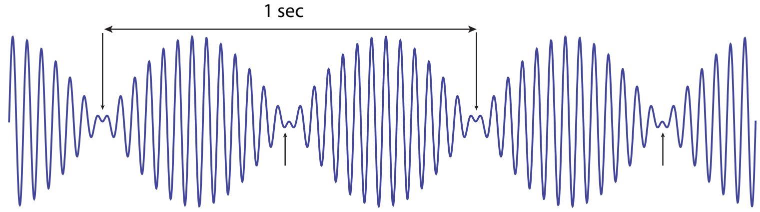


Figure 24.4
A 27 Hz sinusoid and a 25 Hz sinusoid are added, and show beating at 2 Hz, but periodicity at 1 Hz.

periodicity of the tone. However, we expect to hear a 2 Hz beating, since the partials differ by 2 Hz, and we do. In fact, there are two *loudness maxima* per second, but as close inspection of figure 24.4 reveals, *adjacent maxima are very slightly different*. We can't hear that slight difference; rather, we hear the maxima. Mathematically, this happens because the cosine oscillates at half the difference in frequency, or 1 Hz. However, there are loudness maxima at 2 Hz, because they happen whether cosine is near 1 or −1.

Generalizing the Concept of Beats

There are two ways to extend the notion of beats. One is to keep the percentage change in frequency small between the two sinusoids but raise the base frequency so high that the beats repeat with a frequency in the audio range. An example: add 3000 Hz and 3080 Hz sinusoids; the 80 Hz beating may produce an 80 Hz "phantom" beat tone that has no basis in any partials that are present. We take up the topic of beat tones in chapter 25.

Another extension of the idea of beating is to use sinusoids that differ slightly not from each other, but rather ones that differ slightly from a musical interval such as an octave or a fifth. Like the string with its naturally mistuned harmonics, this may give rise to periodic oscillations of timbre—*timbre beating*—with the beating slow enough to count. The changes in timbre are usually small, as befits the reasons for phase insensitivity mentioned above.

24.3 Waveform Beats and the Phase Law

We have already seen a type of *waveform beat* in figure 24.4: a slow cycling of the shape and size of the individual oscillations. The two frequencies 25:27 were quite close, so we call this *1:1 beating*. The waveform oscillations are evident in figure 24.4 as undulations of the envelope; these undulations are also plain to hear.

Here, we discuss a more subtle form of beating, wherein the two partials are quite different in frequency, but may be approximately related by frequency intervals like 1:2, 2:3, 4:5, and so on. The waveform of the addition of two sinusoids related by an octave, a perfect fifth (3:2), a perfect fourth (4:3), and so on depends on the relative phase of the two partials. According to the Ohm-Helmholtz phase law, we are insensitive to this phase.

A closely mistuned partial, say, off by 0.25 Hz, as in 100.25 and 200 Hz, *can be considered to be a perfectly tuned partial with its phase drifting by*

2π *once every 4 seconds* (see equation 24.3). Such a slow drift would allow the listener to establish the timbre at all times, and if the timbre depends slightly on the relative phase, the timbre should cycle through a change once per 4 seconds. This would be interpreted as a sensation of slow, weak beating, not so dramatic as loudness beating. In fact, beats are heard going through a complete cycle once every 2 seconds—see the following. This phenomenon has generated lengthy discussions as to its cause and a prodigious amount of experimentation. A plot of the waveform over some interval (say, 10 seconds) reveals that it cycles through "shapes" at the same frequency as the beats that are heard. This phenomenon has earned the names *waveform beats*, *beats of mistuned consonances*, and *quality beats*. For example, 201 Hz added to a 400 Hz partial beats at 2 Hz; a 200 Hz plus a 401 Hz partial beat at 1 Hz. The reader should try this, using, for example, MAX *Partials*. The beating is subtle but definitely present, even at low volume. The beating may be more pronounced if the upper partial is less loud than the lower one.

We notice that both 201 + 400 Hz and 200 + 401 Hz differ from perfection by 1 Hz, but the waveform beating is different. The waveform *shape* is the way two sinusoids of different period are combining: Are crests adding with crests, and the like? A key point is that a given crest in a plot of $\sin(ax + b)$ shifts at a rate inversely proportional to a as b changes: $\Delta x_c/\Delta b = -1/a$, where x_c is the position of a crest. So, for example, as the phase $b = 2\pi\,\Delta f\, t$ advances in the term $\sin(2\pi f t + 2\pi t)$—($\Delta f = 1$)—it advances the sinusoidal peaks in proportion to $1/f$. This is why 200 + 401 Hz has waveform beats half as often as 201 + 400 Hz.

Some waveform beats are seen in figure 24.5. Although all four examples are strictly periodic at exactly 1 Hz, they regain shape at different frequencies, 1 Hz for 100+201, 2 Hz for 101+200, 2 Hz for 200+301, and last 3 Hz for 201 + 300. The beating we hear makes perfect sense if we are slightly sensitive to the waveform.

We don't need to use whole numbers. For example, beating at 2.5342… Hz is heard for the combination 101.2671 . . .+200 Hz. The actual waveform may never exactly repeat, but the waveform *shape beats* repeat reliably at 2.5342… Hz.

Figure 24.6 shows some detail in a 1-second trace of the waveform $\sin(2\pi \cdot 200t) + \sin(2\pi \cdot 301t)$; the shape changes in the waveform are clear. In spite of near-coincidences, the exact function (as opposed to its overall shape) does not repeat in less than 1 second. The two red traces differ in time by 1/2 second and have the same shape, although they are not quite identical in detail. The red and black traces at the bottom are the same shape except they are each other's negative (one is upside-down compared to the other). They are only a quarter-second apart. If they sounded the same, the beating would be 4 Hz, but instead it is 2 Hz. Thus a waveform and its negative *generally do not sound exactly the same*, even though they differ only by an overall sign. The *relative* phases of the partials are the

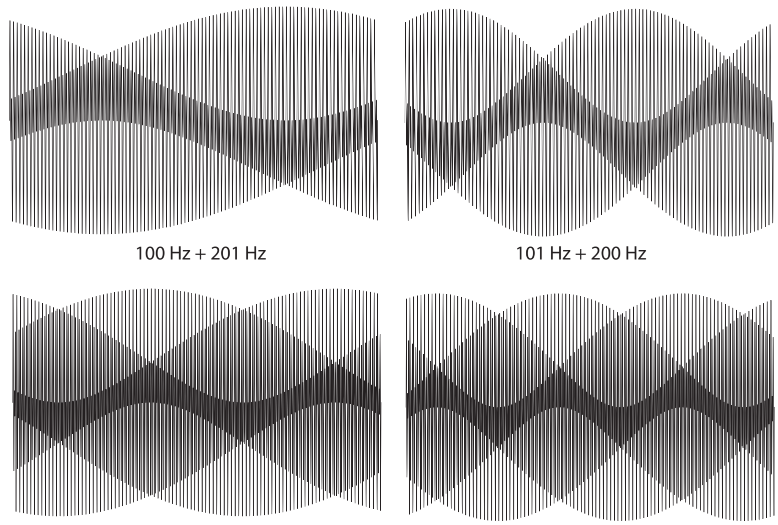


Figure 24.5
One-second sound traces showing waveform beating at various frequencies for combinations of pairs of partials that are all strictly periodic with a period of 1 Hz. We hear mild beating at the frequency of the waveform repetition. (Upper left) Beats repeat once per second. (Upper right) Beats repeat twice per second. (Lower left) Beats repeat twice per second. (Lower right) Beats repeat three times per second. The sound trace repeats only once per second in each case, despite appearances.

same whether or not the signal is inverted; so now we know that even an overall phase (of π or 180 degrees) of the whole waveform can change the timbre. Apparently, we *can* hear the sign of a waveform, which is easy enough to check, assuming one's sound reproduction equipment is responding linearly. The author finds noticeable differences between the sound of waveforms 1 and 2 of figure 24.7.

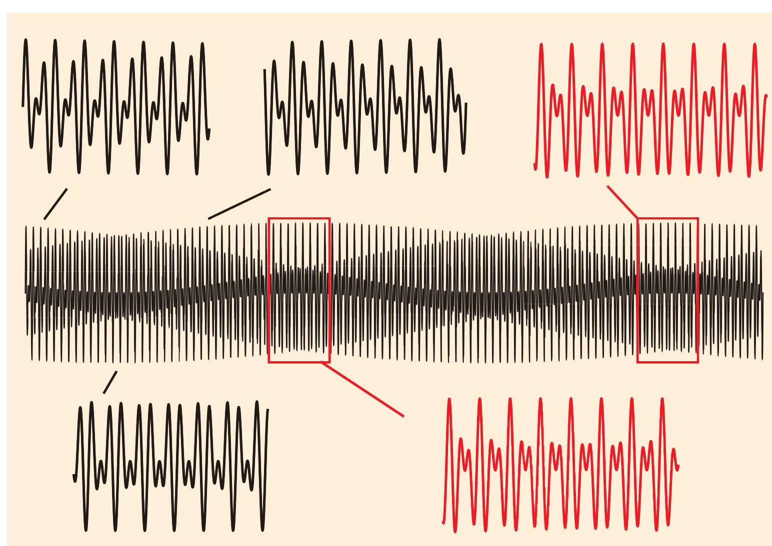


Figure 24.6
A 1-second trace of the waveform $\sin(2\pi \cdot 200t) + \sin(2\pi \cdot 301t)$; insets show details of the waveform at the times indicated. The waveform "shape" undergoes two complete cycles in one second—that is, a 2 Hz waveform beating. The two red boxes and associated waveforms differ by 1/2 s. The exact period is 1 s, so although they are the same shape, they are not quite identical, despite appearances. For the purposes of waveform beating, they sound identical.

Suppose a waveform is periodic but not symmetrical in time—that is, it looks different if reversed (played backward), which corresponds to setting t to $-t$ in the Fourier series. Both original and reversed waveforms have the same spectral content and differ only in phases:

$$
\begin{aligned}
& a_1 \cos(2\pi f t + \phi_1) + a_2 \cos(4\pi f t + \phi_2) + \cdots \\
\to\ & a_1 \cos(-2\pi f t + \phi_1) + a_2 \cos(-4\pi f t + \phi_2) + \cdots \text{(reversed)} \\
=\ & a_1 \cos(2\pi f t - \phi_1) + a_2 \cos(4\pi f t - \phi_2) + \cdots, \qquad (24.5)
\end{aligned}
$$

since $\cos(x) = \cos(-x)$. Depending on the waveform, the sound and the reversed sound are also slightly different in timbre.

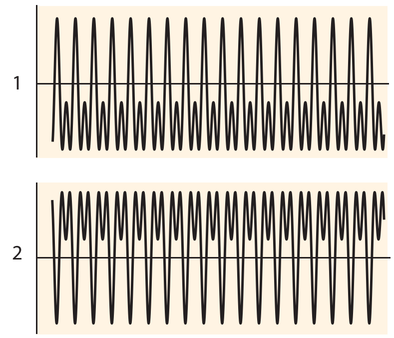


Figure 24.7
The two waveforms 1 and 2 differ only in an overall sign. However, they sound somewhat different—they have a slightly different timbre.

24.4 The Perception of Waveform Beats

The waveform beating debate began with Johann Scheibler (1777–1837), a silk merchant in Crefeld, Germany, who did some early experiments with tuning forks in the 1830s. In 1881, one of the eminent acoustical researchers of that time, R.H.M. Bosanquet, in an article in the *Journal of the Royal Musical Association* titled “On the Beats of Mistuned Consonances” said of his quest to find the source of the beating:

> It is hard to believe that a question to which the answer is tolerably simple could be so difficult. Yet it is very difficult; it is one of the most difficult things I ever tried to do.

Arguments and experiments continued with Ohm, Koenig, Helmholtz, and many others. If humans cannot hear the phases at all (which, we have already seen, is not the case), then some other explanation of waveform beating is needed. The 2 Hz beating sensation for the combination 101 + 200 Hz is consistent with a nonlinear aural harmonic of $2 \times 101 = 202$ Hz generated by the ear itself, beating in the 1:1 way at 2 Hz with the “real” 200 Hz partial. The plot thickens, however, when we try to explain the 3 Hz beats heard when 201 Hz and 300 Hz partials are present. The third harmonic of 201 and the second of 300 are 3 Hz apart and could cause 3 Hz beats, but this is starting to feel like a nonlinear conspiracy theory. *Remarkably, the beating is heard even if one partial is fed to the right ear and the other to the left.* This eliminates some types of physical nonlinear effects as responsible for the beating.

The waveform for adding 200 and 301 Hz sinusoids is cycling twice per second (see figure 24.5); 2 Hz beating is heard. By very clever masking experiments, Plomp[2] showed convincingly that nonlinear distortion is not the cause of the 2 Hz beating. According to Helmholtz's nonlinear ideas, a 2 Hz beating could be, for example, between a 99 Hz nonlinear distortion product $2 \times 200 - 301 = 99$ with a 101 Hz distortion product $301 - 200 = 101$. Plomp masked the region around 100 Hz with noise, which failed to kill the beating, strongly arguing for waveform beating as the mechanism.

Life gets a lot simpler if we merely acknowledge that the Ohm-Helmholtz phase law is only approximately true. This law is not fundamental physics but an observation about human perception, a perceptual trait we just decided is an advantage (see figure 24.3). Suppose that the sensitivity to phase is merely *weak* instead of *nonexistent*, growing weaker still at high frequencies (above about 1500 Hz). At low frequencies, some sensitivity to phase can be assigned to the need for temporal resolution of events. Since it is not pitch that changes in waveform beating, nor loudness, we must assign the beating to periodic changes in timbre. We conclude that timbre can be slightly sensitive to phase. The question then becomes, *how* is it that we are slightly sensitive to phase? We must somehow be sensitive to the *shape* of the arriving waveform.

To summarize, if we humans detect exactly the frequencies that are present and no more (strict place theory), there should be no waveform beats. Timing theory suggests that we might be able to hear differences in the waveform even if the power spectrum is the same. As the relative phase of the partials (and nothing else) changes, the waveform changes, possibly drastically. If beats are heard, we can hear the shape of the waveform, contradicting the Ohm-Helmholtz phase law.

24.5 A Dramatic Phase Sensitivity

A pitched pulse waveform created in the MAX patch *Partials* highlights phase sensitivity and the limitations of the Ohm-Helmholtz phase law, as seen in figure 24.8. Pitched pulses can be created by giving amplitude to a group of partials peaked at frequency near f_{peak}, well above the lowest partial (base frequency) f_0. The pitch of the pulses is centered at f_{peak}, but the pulses repeat with the base frequency f_0. The pulses require a particular phasing of the partials, as is made clear next. Three versions were created, one with phases all the same (which create the

[2] R. Plomp, "Beats of Mistuned Consonances," *Journal of the Acoustical Society of America* v. 42 (1967), 462.

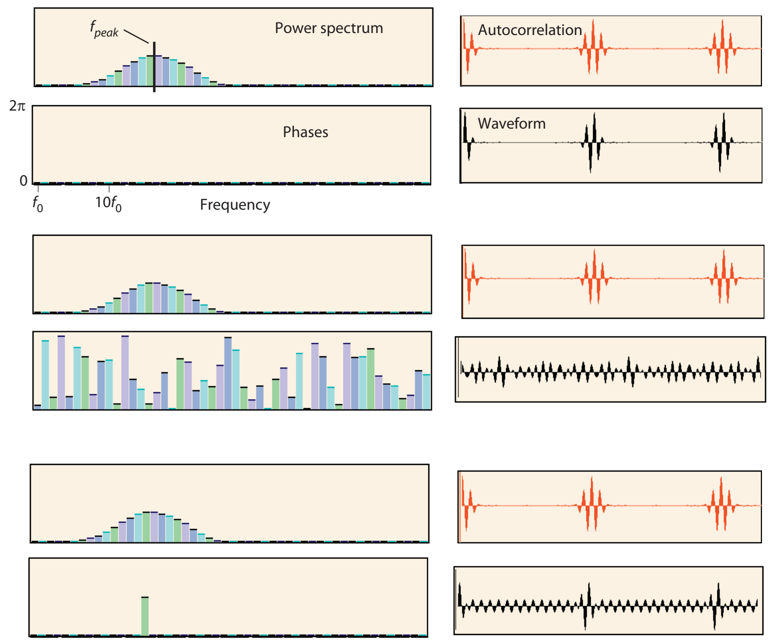


Figure 24.8

Experiment in MAX *Partials* showing an example of extreme sensitivity to relative phases of the partials, in contradiction to the Ohm-Helmholtz law. The autocorrelation function is independent of the phases and is a sequence of pitched pulses. The waveform, however, is very dependent on the phases. When they are randomized (middle example), a disorganized waveform results. When only one partial is phase-shifted, the delicate phasing to make the pulses is upset and the corresponding sinusoid stands out, not only on the waveform but also to the ear, marking a large phase sensitivity. The random phase case also sounds quite different, if the fundamental frequency is below a few hundred Hz. Above that, the differences between the different phasings start to diminish.

pitched pulses), one with random phases (which gives a disorganized waveform but the same autocorrelation), and the last with only one phase altered (the rest again the same), which again gives pitched pulses, except a sinusoid belonging to the altered phase partial (the partial with the phase shift) stands out at all times. The sound changes radically according to the relative phases of the partials, for any base frequency f_0, from 5 Hz to hundreds of Hz. Clearly, the Ohm-Helmholtz law fails miserably—the sound changes a great deal with phase changes in this example.

24.6 Timbre and Context

A bell struck in the normal way can have a luscious timbre. Yet, if the continuous bell tone is recorded, and then played back with an abrupt beginning, or perhaps a smooth onset, it sounds very clinical.[3] The same can be said for many instruments deprived of their normal attack.

[3] Listen to BellSegment.wav on whyyouhearwhatyouhear.com.

Box 24.1

Helmholtz's and Koenig's Ingenious Tests of the Ohm-Helmholtz Phase Law

Ever the resourceful experimentalist, Rudolf Koenig set out to test sensitivity to the phase of the partials in periodic tones. It might seem sufficient just to rotate different sets of holes in a siren relative to each other to change the phase, but this also affects the *amplitude*, not just the phase, of any partials shared by the two tones. For example, a circle of 20 holes and a circle of 30 holes create a 200 Hz and a 300 Hz tone if the disk rotates at 10 revolutions per second. The two tones, which are a perfect fifth apart, both have amplitude in a 600 Hz partial. If both partials had the same amplitude and phase, the amplitude of the sum would be twice that of each of its components. But a 180-degree phase change of one of them would change the sign, canceling the amplitude of the 600 Hz partial. Since a change of phase of one of the notes changes the amplitude of the 600 Hz partial, the timbre will change, since it depends on the relative intensity of the partials.

Helmholtz's circumvention of this problem led him to invent the double siren (figure 24.9). This ingenious device consisted of two independent sirens strongly coupled to tunable Helmholtz resonators. The resonators were supposed to filter out all the harmonics of the siren but one each, selected by adjusting the air volume of the brass chamber. The relative phase was adjustable by changing the timing of the puffs of air. The two siren disks rotated on a common shaft, guaranteeing that the frequencies produced by the two sirens would be locked into simple integer ratios depending only on the number of holes being played in a given disk. The objection to this setup is that the filtering would not have been perfect, and weak harmonics could reinforce and cancel as described earlier, causing unintended subtle alterations in timbre and loudness.

Realizing this, Koenig invented his *wavetable synthesis* method, or *wave siren*. A slit of air emerging from a pressure source strikes a proximate rotating metal band of variable height. The air escapes in proportion to the height of the unobstructed portion of the slit (figure 24.10). The shape of the bands were painstakingly computed and cut, using a photographic process to reduce a large mockup of the curve to the right size for the template. However, the vagaries of real airflow around such obstacles may have caused far more deviation from a pure phase effect than any error in the curves.

He used this apparatus to successfully investigate phantom beat tones, discussed later, but it was not clear that the method of blowing air across cut metal bands as in figure 24.10 produces tones with partials exactly of the same amplitude from one set of phases to the next.

These difficulties led Koenig to construct a real masterpiece, in which he simplified the waves to a sinusoidal shape, and returned to rotating

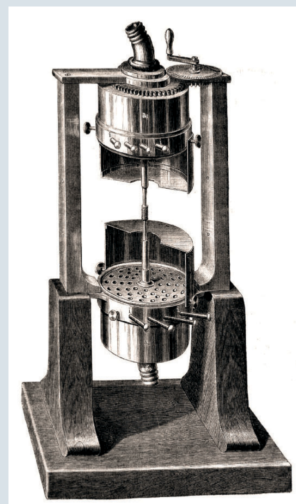


Figure 24.9

Helmholtz's double siren (constructed by Koenig), invented to test senistivity of our hearing to the relative phase of the partials. Each siren drives its own Helmholtz resonator, with a controllable phase.

Perhaps the most dramatic demonstration of context on timbre is to play a note backwards. A piano works especially well. As Murray Campbell and Clive Greated state in their book, *The Musician's Guide to Acoustics*, if a piano is played backward the "instrument is transformed into a leaky old harmonium, although only the order of presentation of the sound has changed." Indeed, the timbre as

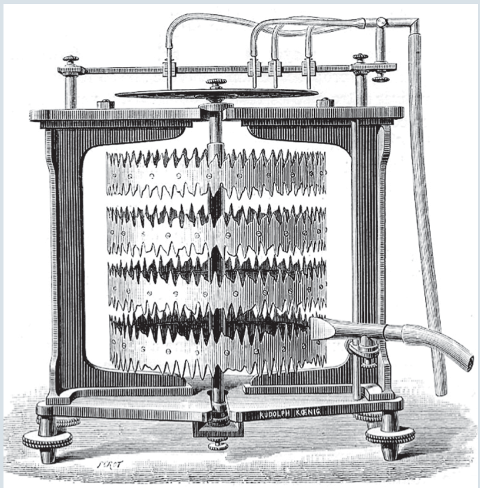


Figure 24.10

The wavetable synthesis apparatus. In this innovative variation on a disk with holes in it, Koenig achieved something close to arbitrary waveform generation, with metal bands cut to the shapes of different waveforms. Using a slit source of air, he could get the pressure variations at the generator to mimic the waveform. In one set of experiments, he used bands that differed only in the relative phases of their partials. By this mechanism, Koenig heard significant differences between different phasing of the periodic waveforms produced by this apparatus. However, given the complexities of the airflow past the metal bands, it is not clear that the test is purely a difference of phase, as his critics pointed out at the time. Another set of bands (the one shown) consisted of the superposition of two Fourier components or partials, which he used to demonstrate beat tones. The bottom band reveals a clear beat pattern from adding two sinusoids of nearby frequency. There is a standard disk siren mounted horizontally at the top for producing various siren tones for comparison.

disks (15 of them), each with double the number of sinusoidal oscillations as the one before, so that he had total amplitude and phase control of 15 harmonic partials, an amazing feat for its day (figure 24.11). Still, Koenig could not be sure that the disks produced only a pure sinusoid.

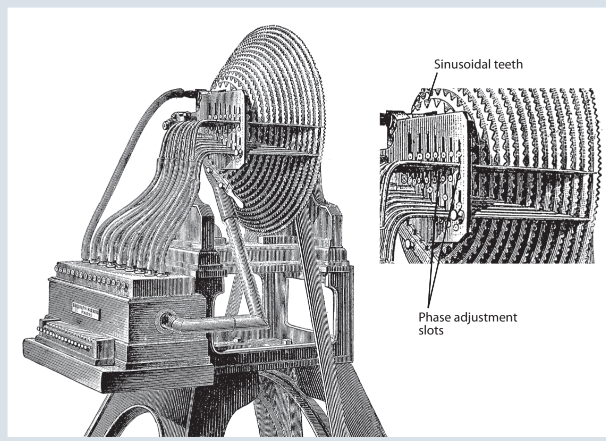


Figure 24.11

This 15-disk wave siren provided complete amplitude control (by varying the air pressure in individual tubes) and phase control (by adjusting the position of the air tubes) of sound.

defined by the power spectrum has not changed, but the perceived timbre goes from a lovely grand piano to something rather unpleasant.[4]

[4]Listen to GrandPiano.wav, on whyyouhearwhatyouhear.com.

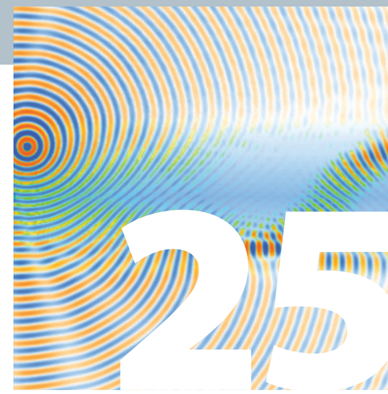
Sound that ramps up and suddenly ends can be turned into pulses that start abruptly and decay slowly, just by reversing them. This does not affect their spectral content, but they leave a very different impression. This is an easy experiment to try on your computer.

24.2 Timbre, Loudness, and Shock Waves

Universally, when wind instruments (including the voice) are driven harder, the strength of the higher harmonics grow relative to the lower harmonics. The sound becomes more brilliant, or perhaps develops too many high harmonics and begins to sound raspy, according to the effects of autodissonance and overlap of harmonics on the basilar membrane. The vocal folds, for example, suffer more violent and abrupt opening and closings when driven at higher pressures, which necessarily generates stronger high harmonics.

An interesting phenomenon happens with both the trombone and the trumpet, and possibly other wind instruments: the oscillating air column vibrations arrange themselves into a shock wave under very loud driving of the instrument; this has been captured using schlieren photography. The shock wave certainly requires the presence of high harmonics. Perhaps most surprising, however, is that a sharp shock front implies a precise phasing of the harmonics of the air column, in analogy to the Helmholtz wave on a violin string.

How indeed are the relative phases of the harmonics determined under any playing conditions? Certainly they are not random, since the vocal folds (or lips, in the case of a trumpet, for example) are open in pulses, and the pressure due to a given partial should be high in the mouthpiece at the moment of the pulse to resonantly enhance that partial. This suggests a more pulsed waveform than a random choice of phases is likely to produce. It appears that further investigation of the relative phases in wind instruments, as a function of the player, instrument, lipping up or down, and so on would be very rewarding.



Phantom Tones

Reality is merely an illusion, although a very persistent one.

—Albert Einstein

In this chapter, we discuss perceptions of tones that simply are physically not present in the sound arriving at out ears. Certainly, we see things that aren't there, and many good visual illusions are widely known. Auditory illusions are much less well known, but we try to partially remedy that here.

25.1 Lies and Illusions

Illusions are of two types: (1) those with a direct purpose, which we like to call "lying in order to tell the truth," and (2) just plain weird and "unexpected" side effects of our sensory apparatus and algorithms.

An example of the first type from the visual world is two squares on an image of a chessboard that look like they are very different shades of gray (A and B in figure 25.1), when they are in fact physically the same shade of gray on the printed page.[1] This example falls under "lying to tell the truth," since it is very likely that the chessboard would have been uniform under uniform illumination, and our brains know that the shadow of the cylinder should cause only an apparent, not a real darkening of the shaded region. This is how the image is presented to our consciousness, which is a lie because the areas that appear to be much darker in shade are not.

A very clear audio illusion with a definite purpose involves estimating the direction of a sound source. A sudden pulse of sound from the right will arrive at the right ear first. We use such arrival time delays to help decode

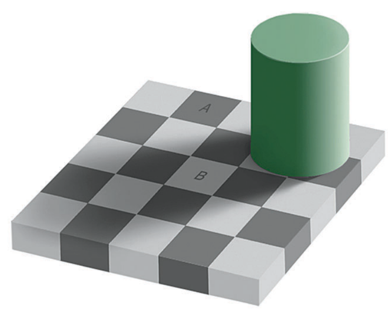


Figure 25.1
A visual example of our sensory system and the brain lying to us in order to tell the truth. The two squares of the chessboard, labeled A and B, are of exactly equal gray value. Courtesy Edward Adelson, MIT.

[1]The image is also found on whyyouhearwhayyouhear.com. Copy it to your screen and experiment with it—for example, cut and paste the two regions in question onto a blank field.

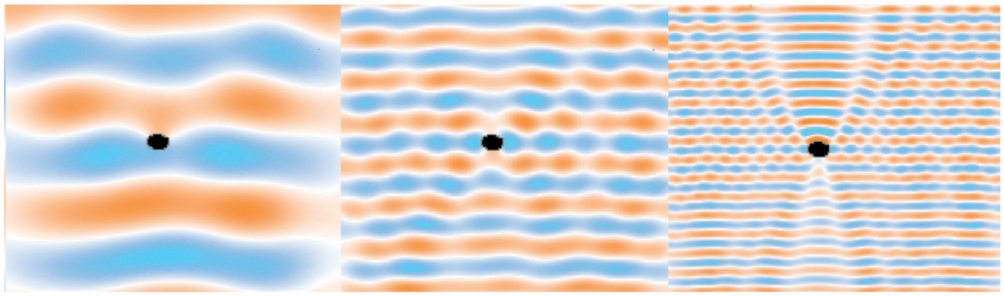


Figure 25.2
The sound field near a model head at three different frequencies, 220 Hz (left), 600 Hz (middle), and 1400 Hz (right). The incident sound is a plane wave coming from above. Note that the sound intensity is about the same on the side of the head facing the source as it is on the side facing away at 220 and 600 Hz, but it starts to diminish on the far side as the wavelength approaches the size of the head. So, for low to midrange frequencies, the sound is nearly equally loud on either side of the head, but there is a crucial time delay, so that both ears are not receiving the same signal at the same time. The impression that the sound is much louder in one ear than the other is a necessary illusion, designed to quickly reveal the direction of the source of the sound.

where the source is, as mentioned in section 21.2. A click from the right side of the head is heard *only in the right ear*, yet the sound is almost exactly as loud in the left ear after it has diffracted around the head! Our brain suppresses the sound on the left, which is a lie with a purpose: to convey, without delay, the impression that the source of sound is to the right. How else would this information be presented to us so that the conclusion is instantaneously obvious?

The diffraction of sound around small objects can be simulated in *Ripple*. Draw a mock head receiving sound from one side, as shown in figure 25.2; make sure that the Fixed Edges option is unchecked. With two probes and a source, show to your satisfaction that (1) the sound is almost as loud at the "far" ear if the wavelength is long enough, and (2) the phase of the arriving sound differs compared to the "near" ear.

This illusion is easy to quantify using earbuds and sound generation software. Use your laptop to generate or record a sharp click in monaural sound. Copy it over to a second stereo channel, and then time-delay the playback in the right ear by various amounts. You can do this by using the Generate Silence option in one channel after copying the click over. Using earphones, try time delays of a quarter of a millisecond up to a few milliseconds. For delays of about 0.66 millisecond, which is the time delay for sound to cross the distance spanned by a human head (the so-called interaural time difference), you will perceive that the sound is coming from the side with the first arrival of the sound. but more than that, even though you know the intensity of the sound is the same in both ears, *it will sound much louder in the ear with the first arrival.* (High frequencies are more shadowed by the head and do result in interaural level differences, also used as cues for sound localization.)

A visual illusion of the second type, a side effect that probably brings no advantage, is seen in figure 25.3, where light gray spots appear in the intersections of black stripes against a white surround. This is an untruth that results from some no doubt very useful visual algorithms, leading to distortions of reality with no purpose in special circumstances.

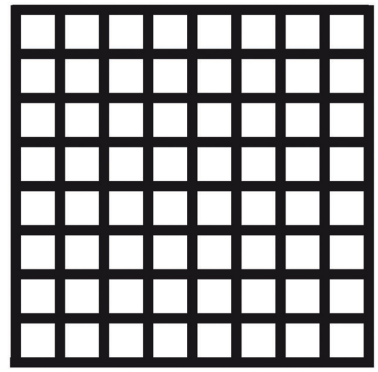


Figure 25.3
If you stare at the intersection of two of the black lines, you will see lighter gray spots at the adjacent intersections, a clear visual illusion that is a side effect of our visual processing algorithms.

25.2 Sounds That Aren't There

There are several phenomena related to tones that are perceived but not present. Perhaps they are the analog of the visual effect just mentioned: side effects of auditory processing. They go under the names of Tartini tones, difference tones, combination tones, beat tones, resultant tones, distortion products, differential tones, summation tones, and no doubt more. Some of these terms are overlapping, and others have not ever received crisp definitions or usage. We will not succeed in defining a zoo of these effects with every one in a different cage.

A *difference tone* is perceived at the difference frequency $f = f_2 - f_1$; a *summation tone* at the sum frequency $f = f_2 + f_1$. *Tartini tones* are generally applied to compound (periodic but not sinusoidal) generators, whereas difference tones refer to simple sinusoidal generators. We need a blanket term, one that acknowledges that all these effects, in spite of their nuances, do share some common roots. We call them *phantom tones*. A phantom tone is a tone not in the sound presented to the listener but heard by the listener nonetheless. Are all the various phantom tones essentially the same phenomenon?

We have made the case that pitch is not intrinsic to sound and is a human sensation, so is it not phantom? Pitch and phantom tones are independent phenomena, because pitch can exist without the perception of any *tone* at the frequency of the pitch. For example, a good chime has a well-defined pitch, yet no partial at that pitch is heard and no sensation of a tone is present at that frequency. A phantom tone, when it is present, sounds as real as an instrument playing that note, if faintly.

Hearing Phantom Tones

Combinations of real tones (*generating tones*) may spawn the sensation of other tones or partials that aren't physically present. Phantom tones sound perfectly real—tones with a pitch—but their perceived strength is dependent on the listener, context, and training. It is possible to draw attention to phantom tones by causing them to change pitch, which can be done by varying the pitch of the generating tones. Sometimes alternately removing and restoring a generating tone brings out a phantom tone. These techniques are analogous to the trick used to make partials stand out—both cause us to switch from synthetic to analytic listening. Practice makes it possible to make the switch without such coaxing.

What relation does the frequency of a phantom tone bear to the generating tones? What is the connection between fast beating and phantom tones?

A way to discover new phantom tones is to use two pure sinusoidal partials as the generating tones, and then listen for tones at other frequencies. Some sine tone generating software allows slowly ramping up one partial in frequency, while keeping another fixed. This can be done using the cursor control to ramp a partial up or down in frequency in MAX *Partials*. Some people seem to have difficulty recognizing phantom tones, possibly because they have more accurate hearing or are less capable of analytic listening—it is difficult to tell which. There is growing evidence for a considerable variability in sound processing in the brain from person to person. A dramatic example of this is Deutsch's audio illusion, discussed in section 25.21, which different people hear *radically* differently. For this reason, there may be no fixed set of answers to the question of what phantom tones are heard under a given set of circumstances.

A key question is whether the phantom tones grow louder in proportion to the generator loudness, or if they respond more dramatically, suggesting nonlinear effects. You can check this on yourself once you find good examples; specific cases to try of frequencies f_1 and f_2 are given in the following.

25.3 How and Where Do Phantom Tones Arise?

The debate on the cause of timbre beating is recapitulated in the debate on the causes of phantom tones.

Mechanical Causes

Camps divide according to what part of the auditory chain is responsible for phantom tones. Place theory maintains that the required partials, absent in the incident sound, are created by the mechanics of the ear, so that the missing vibrations are actually real by the time they are detected by the neural system. This is quite possible on general nonlinear mechanical principles. The presence of two frequencies may in fact generate a third frequency or even several new frequencies, which could be a difference of the two, sum of the two, twice one minus the other, and so on.

Helmholtz was attracted to the idea of nonlinear sound generation in the ear. His thinking was that if we are frequency analyzers, we cannot hear a particular frequency or a tone unless it is physically present. Faced with the absence of the first partial in some complex musical tones—the residue pitch effect—Helmholtz felt compelled to restore the lowest partial, to require its physical presence, by saying that it was produced by nonlinear effects based on the presence of the higher partials. This idea is flawed, since the residue pitch is not actually accompanied by the ability to hear out a sinusoidal partial at the frequency of the pitch. In other words, Helmholtz's phantom partial is not even perceptually present. He was quite sloppy about the distinction between *tone, pitch*, and *partial*, using the words interchangeably, when more precision of language was called for. His translator complained bitterly about this, as we remarked in section 23.9.

Perhaps his mind was clouded by the beauty of the nonlinear idea—if only it were true! It fits a very pretty physical phenomenon that is nearly universal for physical systems that vibrate: nonlinear generation of harmonics and combination tones. Due to this effect, a vibrating system can be driven at one frequency and generate other frequencies spontaneously. *If the vibrational amplitude is large enough*, new frequencies are generated that are combinations and differences of multiples of the frequencies that were originally present. Thus, if we force a nonlinear system with sinusoidal frequencies f_1 and f_2, we might see, for example, a new sinusoidal frequency $f = f_2 - f_1$ generated as a response. In Helmholtz's day, this notion was new, and it must have been tempting to appeal to this mechanism in the face of the apparent dominance of the fundamental. It assigns the perception of pitch at a missing fundamental frequency to a physical property of the ear, amenable to analysis in terms of relatively simple nonlinear oscillations. Still, it is surprising that Helmholtz took this physical, causative path to explaining pitch, since he was comfortable with psychophysical phenomena. He had already spent much time juggling such issues in connection with color vision and sight.

Neural Causes and the Auditory Cortex

Timing theory supposes that phantom tone generation lies further up the neural chain: the nervous system can create, either deliberately or as a side effect of its algorithms, the sensation of tones that reveal repetitious *events* in the sound, present even if the corresponding partials of the same period are absent.

The auditory cortex stands between the ear and the seat of consciousness. It is divided into three parts with different function. The *tertiary* *cortex* apparently synthesizes the aural experience before sending it on, and the *secondary cortex* apparently processes harmonic, melodic, and rhythmic patterns. The *primary cortex* extracts pitch and loudness data, and is *tonotopically* organized (different frequency zones in physically different places). Brain scans show that the primary cortex is not involved when imagining music, but it is active when schizophrenics have auditory hallucinations.

Hallucinations

Hallucinations, either visual or auditory, can be indistinguishable from reality. One of the maladies that Oliver Sacks recounts in his fascinating book *Musicophilia* can be distracting to the point of despair: people quite suddenly hear music that isn't there, and it doesn't tend to go away. The sensation is nothing like the tune you can't get out of your head. The music sounds completely real. The genre of the music heard is often not a match you would imagine for the patient, who may or may not be musically inclined. If, for whatever reason and through whatever mechanism, the auditory cortex decides to create and send phony data, there is apparently no way for our conscious minds to tell it is not real.

What has this anomaly got to do with pitch perception or phantom tones? The point is that if the auditory cortex can send symphonies that aren't there, why couldn't it more routinely send us the sensation of a "real" tone, which might sound like a single partial or perhaps a complex tone, and which isn't physically present but still represents some aspect of the real sound being processed? Sending such a sensation might serve a purpose or it might be a side effect of complex processing algorithms.

Otoacoustic Emissions

It is known that the ear emits sounds as well as receives them. Signals from the nervous system are sent to the cochlea, causing hair cells to contract and relax at audio frequencies, resulting in sound emission from the ear. (These otoacoustic emissions were discussed in section 21.4.) The relevance for the present discussion is that otoacoustic combination tones (called *upper beat tones* by Koenig, Zahm, and others) of the form $f = 2f_2 - f_1$ are quite discernible by the sensitive microphones used to detect otoacoustic emissions. Thus the combination tones are "really there" in a physical sense within the ear, but it is still not completely clear how they get produced, or whether these signals are the ones we hear, and where in the neurological chain the signals originate.

Helmholtz was wrong about hair cells being little high-Q resonators on their own, but neural feedback effectively makes them capable of sharp frequency resolution anyway. It seems likely that Helmholtz was also wrong about the importance of *mechanical* nonlinear effects, but this too may be rescued by nonlinear neural feedback—that is, otoacoustic emissions.

25.4 Beat Tones

Loudness beats are recognized by a periodic waxing and waning of the amplitude of the resultant wave.

Phantom Loudness Beat Tones

In figure 24.4, we saw the addition of 25 Hz and 27 Hz sinusoids of equal amplitude. (We use smaller frequencies for clarity in plotting.) The GCD of 25 Hz and 27 Hz is 1 Hz, but there are *two* beats per second, if by *beats* we mean broad maxima in the envelope of the higher frequency oscillations. As inspection of the waveforms reveals (see figure 24.4), the two beats each second are not quite the same: only every other beat is an exact repeat, giving a strict periodicity of 1 Hz. Nonetheless, it makes little difference whether the waveform is formally periodic at 1 Hz; we hear 2 Hz beating since the sound is louder twice per second.

A Tone at the Beat Frequency?

The beats at the difference frequency are "events" in their own right, even though there are only two frequencies present. Certainly, we hear those events if they are slow—as in 200 Hz plus 202 Hz—as loudness beats at 2 Hz. As they get faster, at what point would their presence be *completely* inaudible? The point is, there would be some residue of the countable beats—they would, in fact, remain audible. If the difference frequency is in the audio range, we hear this periodic sequence of events as a tone. It may be serving a purpose, to inform us of events at that frequency.

A London police whistle makes just such use of beating of two nearby high-frequency tones. Suppose the whistle generates a 3000 Hz and a 3080 Hz tone. The combination beats at 80 Hz, well into our hearing range. The GCD of 1000 and 1080 is 40, which is the frequency of this combination, also well into our hearing range. The beating tone, however, is heard at 80 Hz, which is the frequency of the waveform pulses. The exact details of each pulse recur only at 40 Hz.

Before Ohm and Helmholtz came along, the issue of phantom tones and beat tones was thought to be resolved. Thomas Young (figure 25.4), an amazingly talented British polymath who helped decode the Rosetta Stone and worked out much of vision theory (and along the way performed an interference experiment with light that is today a paradigm of quantum mechanics), took up the question of phantom tones. The subject was initially raised by the violinist Giuseppe Tartini in 1754 and the German organist and composer Georg Andreas Sorge in 1745. They had heard the beat tones as "third notes" when playing two others. It is hardly credible that it went unnoticed by generations of musicians before. However,

Figure 25.4
Thomas Young (1773–1829). Courtesy Materialscientist.

phenomena are often not attributed to their first discoverer, but rather to someone who described their significance most eloquently. Young promoted the commonly accepted explanation, until Ohm and Helmholtz questioned it.

Young and also Joseph Lagrange first argued that beats are events—that is, loudness peaks owing to constructive interference maxima that occur repeatedly at the beat frequency. The ear is willing to assign a tone to this periodic succession of loudness undulations, despite the complete absence of partials at the frequency of the beat tone. Young said: "The greater the difference in pitch of two sounds the more rapid the beats, till at last, like the distinct puffs of air in the experiments already related they communicate the idea of a continued sound; and this is the fundamental harmonic described by Tartini."

Examples of Beat Tones

Even for just two sinusoids, there are all sorts of cases to consider. It is amazing how many mathematical ramifications there are surrounding the choice of just two numbers. Is their ratio rational or irrational? Are any integer ratios, involving small integers, a good approximation to the ratio of the two? The answers to these questions affect what we hear.

For example, suppose $f_1 = 1000$ Hz, $f_2 = 800$ Hz. These beat at 200 Hz, and perhaps therefore we will hear a 200 Hz tone. But not so fast—200 Hz is also the residue pitch. We should examine instead, say, 1042 and 842 Hz. These are now inharmonic partials, but they differ by 200 Hz. Indeed, a 200 Hz difference tone—a true phantom tone—is clearly heard, weaker than the two generators but still quite distinct.

This sheds some light on residue pitch. Nothing drastic happens to the perceived 200 Hz tone in going continuously from 1042 and 842 as generators to 1000 and 800. This implies that any tone heard along with the 200 Hz pitch is an *event tone*—that is, a tone generated in our sound processing hardware and software to signify the occurrence of repetitive events (beats) at a frequency of 200 Hz in either case.

A case is provided by intervals near 8:15, using frequencies such as $f_2 = 2048$ and $f_1 = 3840$ Hz, giving the beat tone $f = 2f_2 - f_1$ of 256 Hz, which indeed is heard. *This is also just the frequency of the waveform beats.* If f_2 is raised by A Hz, the phantom tone increases by $2A$ Hz, as does the frequency of the waveform beating. If f_1 is raised by A Hz, the tone is *lowered* by A Hz, as is the frequency of the waveform beating. The reader is encouraged to try this, using, if possible, high-quality earbuds and good tone generators. One may vary f_1 by 200 Hz in either direction and f is clearly heard, obeying $f = 2f_2 - f_1$.

Another interval, $f_2 = 2048$ and $f_1 = 3072$ Hz (and analogous intervals), is remarkable in that, first, the difference tone $f = f_1 - f_2 = 1024$ and the combination tone $f = 2f_2 - f_1 = 1024$ are the same; thus what Koenig termed the *lower beat tone* (the *difference tone*) and the *upper beat tone* (the *combination tone*) coincide. This combination has loudness beats at 1024 Hz. If the higher frequency f_1 is raised by A Hz, the difference tone should increase by A Hz, but the combination tone should *decrease* by A Hz. Indeed, as $f_1 = 3072$ is lowered by hand, as is possible in the MAX patch *Partials*, one hears both rising and falling phantom tones (a rising combination tone and a falling difference tone) interfering with each other, causing beats when they are still relatively close in frequency. When $f_1 = 3073$ Hz, there should be two beats per second, and that is what is heard. *This is also the frequency of the waveform beats.* If, on the other hand, the lower frequency is raised by A Hz, the difference tone should decrease by A Hz, but the combination tone should increase by $2A$ Hz, implying a beating of between them of $3A$ Hz. That is heard also. Quite an instructive example! We are feeding our ears only two pure tones, both above 2000 Hz, yet we are discussing the easily audible beating of two phantom tones, both around 1000 Hz!

25.5 Nonlinear Harmonic Generation

The key to hearing interesting phantom tones is to make sure they do not already exist in the sound presented to the listener, owing to some quite common imperfections in the production of sound. In the old days, Helmholtz, Seebeck, Ohm, and especially Koenig went to extraordinary lengths to ensure that there were only pure sinusoids coming from tuning forks, sirens attached to resonators, and the like. Nonetheless, aspersions were sometimes cast regarding the purity of a competing researcher's sound sources. Today, professional-level recording and playback equipment may be employed to ensure the near absence of contaminating frequencies. Laptops and earbuds may not be free of such problems.

When a single vibration mode is present—that is, a single simple oscillator such as a real (as opposed to an ideal) pendulum—an oscillator forced sinusoidally at frequency f may generate other frequencies $2f$, $3f$ ... as well. This is called *harmonic generation*, but it may sound like old news: don't simple strings have harmonics? They do, but the situation is quite different. In a string, each mode $f, 2f, \ldots.$ is an oscillator in its own right—a string has many *different* vibration modes. Also, in real strings the higher modes will not be quite exactly integer multiples of the lowest mode frequency, so that the second partial of a string might be 401.3 Hz if the first partial is 200 Hz. If on the other hand a single mode is oscillating in a periodic but not sinusoidal way, higher harmonics must be present and are *exact* multiples of the lowest frequency. An ideal pendulum or mass and spring oscillates sinusoidally, so just one frequency is present. Harmonic generation is associated with "nonlinear" vibration. We can get a feel for

Box 25.1

Experiment in Nonlinear Harmonic Generation

Figure 25.6 is an instructive case study in nonlinear generation of tones that are not there in the original signal. They were generated by deliberately overdriving cheap analog electronics (earbuds and microphone). The sinusoidal frequencies f_1 and f_2 were generated on a laptop (the frequency of f_2 was ramped up linearly with time) and fed at loud volume to earbuds, one of which was placed very close to the inexpensive microphone of a dictation headset. The digital sound generated by the computer thus passed through analog stages (earbuds, microphone) before becoming digital data again in the computer. The analog processes are subject to harmonic generation and other nonlinear distortions, which become evident in the sonogram of the data.

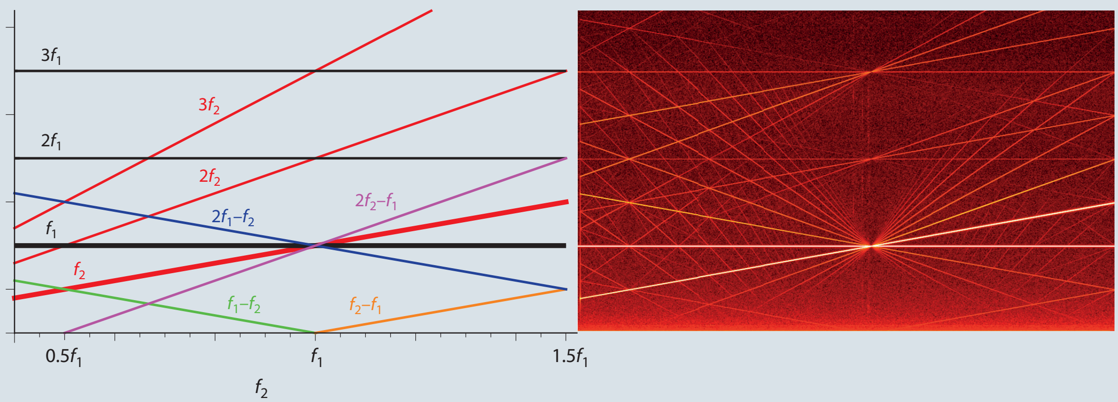


Figure 25.6

(Left) The thick black and red lines represent a fixed sinusoidal partial f_1 and a rising partial f_2. Harmonics and difference frequencies are shown in lighter lines. For frequency f_2 and fixed f_1, several harmonics and difference tone frequencies for the lowest orders are shown. It is seen that sometimes there are coincidences of various orders—that is, where the light lines intersect. For high-pitched f_1 and f_2, the important phantom tones are below, and sometimes well below, f_1 and f_2. (Right) Sonogram obtained as follows: The sine tones f_1 and f_2 were generated on a laptop (the frequency of f_2 was ramped up linearly with time) and fed at loud volume to earbuds, one of which was placed very close to the microphone of a dictation headset. The digital sound generated by the computer thus passed through analog stages (earbuds, microphone) before becoming digital data again in the computer. The analog processes are subject to harmonic generation and other nonlinear distortions, leading to the weaker lines seen corresponding to the tones specified on the left, as well as others not shown on the left. The hardware produces distortion products, which if presented to the ear are real, not phantom. Helmholtz suggested that this kind of distortion happens mechanically in the ear, so that the phantom tones we hear are in fact real by the time they are detected.

Box 25.2

Rudolph Koenig

Rudolph Koenig was a unique figure in acoustics. Koenig did not do well in college (he had trouble with languages) and apparently never wanted to be a part of the academic establishment. He became, nonetheless, a key player in the scientific controversies of his day concerning acoustics and perception. He made a living selling demonstration equipment and scientific instruments. Born in Königsberg, he apprenticed for eight years under the violin maker Jean Baptiste Vuillaume. He began a new career making scientific instruments and a few years later had earned a reputation as perhaps the most talented and brilliant instrument maker of his time, a true artisan. His instruments were exclusively for the purpose of demonstrations of sound and hearing. He was also a creative and talented scientist, involved in the acoustical controversies of his day (and our day, as we have said). He built better and better instruments to answer the key experimental questions, and other innovations meant to illustrate fundamental acoustical principles and his interpretation of the questions surrounding human hearing. Koenig became one of the best researchers in acoustics, at first refining and supporting Helmholtz's ideas. Later, he became Helmholtz's strongest and most effective critic, often using devices of his own design, remarkable considering Helmholtz's stature and the fact that Koenig was not formally educated beyond secondary school.

Koenig had a reputation for meticulous integrity. His reputation for well-conceived experiments and fine instruments was without peer. Although he sometimes put Helmholz's ideas into their finest experimental form, and improved many other instruments, he is best known for making by far the most complete and accurate sets of tuning forks and putting them to remarkable uses. The importance of high-quality tuning forks as standards and investigative instruments is easy to underestimate in a digital world. A Koenig tour de force was displayed at the Philadelphia Exposition of 1876: a "tonometric" apparatus consisting of 670 tuning forks, of different pitches covering four octaves.

Getting the right answers to the subtle questions surrounding the origin of phantom tones requires instruments of high precision, and most precise among these was the tuning fork. The key was to make each fork emit a single pure sinusoid. Even today, it is no mean feat to ensure that loudspeakers and microphone are operating in the purely linear regime, where distortion products play no role. And even today, the gold standard in tuning forks are those made by Koenig.

He enjoyed presenting acoustic and perceptual phenomena to relatively large groups. Scientists, students, musicians, and craftspeople gathered at Koenig's workshop from all over Europe and America. It was a unique place—part home, part commercial space, part institute. They enjoyed Koenig's remarkable scientific and musical demonstrations, part seance and part seminar. One could not disentangle business and science at Koenig's shop. All this took place in the atmosphere of friendliness and the highest traditions of craftsmanship.

Like Seebeck, Koenig took issue with the assertion that pitch is associated with pure partials of the same frequency. Instead, Koenig asserted that pitch and tones sprang from periodicity, which as he knew could exist without a partial of the same period. Seebeck the schoolmaster and Koenig the artisan were right, but their opinion did not carry against the Aristotelian weight of Helmholtz. Even so, it is widely acknowledged that Koenig gave Helmholtz a run for his money on key issues surrounding pitch perception, the nature of difference tones, beat tones, and so on. Some contemporaries even thought Helmholtz lost the arguments. Careful reading and testing of Koenig's arguments and examples are convincing regarding the correctness of his main ideas, if not all the details. This is not to diminish Helmholtz, a truly great physicist. The issue remained controversial for a long time, even up to the present, as such issues do in science when truth and authority do not coincide.

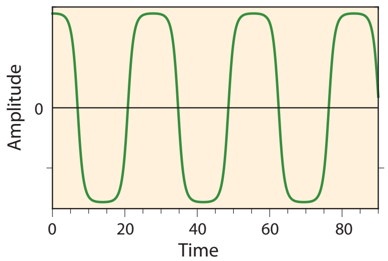


Figure 25.5

Amplitude of a pendulum versus time for nearly vertical initial displacement. The damping was set to zero, and there was no forcing. Notice the distinctly nonsinusoidal oscillation, which through Fourier's theorem will require harmonics of the fundamental frequency in order to reproduce the curve.

generation of new frequencies owing to nonlinear effects by considering a real pendulum at large oscillation amplitudes (figure 25.5). The pendulum was released from rest in a nearly inverted position, so it swings back and forth almost full circle. The pendulum is slow to fall away from the nearly inverted position; this causes a nonsinusoidal shape in the plot of the amplitude versus time. Higher multiples of the fundamental frequency—that is, higher harmonics—will be required to describe this flat-topped curve; something vibrating in this way will emit these higher harmonics as well as the fundamental. If the moving parts of the ear are nonlinear oscillators, they could generate the "aural harmonics" $2f, 3f, \ldots$ if forced at frequency f. Or the harmonics could be caused by neural feedback to the hair cells.