Part D: Characterizing Sound

Chapter 36. Parameters and Perception

Simple Harmonic Motion of an object is itself not sound, but a vibrating object can be a source of sound. Each of the physical parameters of SHM vibration is closely related to a perceptual characteristic of the sound produced.

- The duration of a vibration, how long it lasts, is mainly determines the duration of the resulting sound.
- Vibrations with larger amplitudes create sounds that are **louder,** all other things being equal.
- Vibrations with higher frequencies cause sounds that are higher in **pitch**.
- Since period is the reciprocal of frequency, it also relates directly to **pitch**. A longer period corresponds to a lower pitch.
- The initial phase of a vibration has no bearing on the perception of the sound. Since the value of initial phase depends on a choice made by an observer (specifically, the choice of the time origin), it can't have real-world consequences. Recall from Chapter 21 that phase *differences* can be physically meaningful, but those will only arise when multiple sounds combine.

These relationships are not exclusive, as each of these perceptions (duration, loudness, and pitch) can be influenced by all of the physical parameters except initial phase. For instance, perceived loudness depends slightly on frequency. Nevertheless, the connections above are the principal ones.

The characteristics of a sound can vary over time. Over the course of seconds, like notes in a song, you might consider that separate sounds, or one changing sound. On a more rapid scale, the exact variations over the first few milliseconds of a sound, called the **attack** in some contexts, can be very important for recognizing the source. However, in order to keep things simple, this book focuses primarily on sounds that are uniform throughout their duration.

Some example frequencies will give a sense of the numbers involved. The pitches from a piano have frequencies ranging from about 27 Hz to almost 4200 Hz. Humans are capable of perceiving pure tones roughly between 20 Hz and $20000\text{ Hz} = 20\text{ kHz}$. Part E looks into those limits in more detail. Note that the range narrows as we age, especially at the upper limit, with many adults limited to 15 kHz or lower. Other species are able to perceive more extreme frequencies. At the **infrasound** end (sound below 20 Hz), there are indications that elephants and some whales may hear frequencies of 10 Hz or lower. In **ultrasound** (sound above 20 kHz), many animals (including dogs and cats) can hear frequencies as high as 60 kHz, and bats perceive as high as 100 kHz.

Equivalent examples for loudness are not possible in terms of topics covered so far. Although there is a direct link between amplitude and loudness, there are other factors. For instance, even for a single source (with a particular vibration amplitude), the perceived loudness will depend on how far you are from the source. Numerical measures of loudness are introduced in Chapters 50 and 54.

Specifying a sound's pitch and loudness does not completely describe what it sounds like. In fact, some sounds do not have an identifiable pitch. Even just among musical instruments, the main reason to have different instruments at all is because they sound different from each other. Any characteristic of a sound other than pitch, duration, or loudness contributes to the characteristic called its **timbre** (pronounced 'tam•bər).

While it's simple to say what timbre *isn't,* it's a bit harder to describe what timbre *is.* This is partly because, unlike pitch and loudness, timbre cannot be limited to a single numerical, one-dimensional scale, like more and less or higher and lower. Timbre is sometimes described as the "quality" of a sound, but this is not meant in the sense of better or worse quality. Timbre is described using many different adjectives, some borrowed from the other senses. For instance, it is sometimes described as the color of a sound. Much of Part D is aimed at how quantifiable characteristics can describe timbre.

Chapter 37. Vibration Superposition

Vibrations cause sounds, as with tuning forks and guitar strings, and sounds cause vibrations, as in your ear or a microphone. Section 20c shows that the simplest sources of sound vibrate in a very specific way, simple harmonic motion, to produce pure tones. However, ears and microphones are not restricted in the way that they can vibrate. If you listen to two pure tones at once, both of them are registered in your ear simultaneously, indicating that it can vibrate in a more complicated way. Something that can vibrate in response to sounds is called a **sympathetic vibrator.** If it responds equally well to all different sounds and combinations of sounds, it is a **non-resonant sympathetic vibrator**. The meaning of the adjective “non-resonant” is covered in more detail in Chapters 68 and 69.

How does a non-resonant sympathetic vibrator move, when it responds to multiple sounds? The following is a good model for a great many situations, as long as the sounds are not extremely loud.

> The displacement of a non-resonant sympathetic vibrator in response to multiple sounds is the moment-by-moment sum of all the responses that it would have to the individual sounds. This is called **superposition**, and a vibrator that obeys superposition is said to have a **linear response.**

Superposition is actually a general term in physics, used whenever simply adding things together gives the correct result for their combination. See Chapter 18 for another example. In this case, we are adding curves rather than numbers.

Figure 37.1 shows examples. Part (a) shows the superposition of two pure tones with quite different periods, and it is fairly easy to see how they each have contributed to the combined vibration. At each point along the time axis, the displacement of the combined vibration is the sum of the displacements of the contributing vibrations,

$$\Delta x_{\text{tot}} = \Delta x_1 + \Delta x_2 (+\cdots) \quad , \qquad (37.1)$$

where the parentheses would apply if there were more than two contributing vibrations. In this book, any vibration that is not sinusoidal, such as the results of superposition shown here, will be called a **complex** vibration, in contrast to simple harmonic motion.

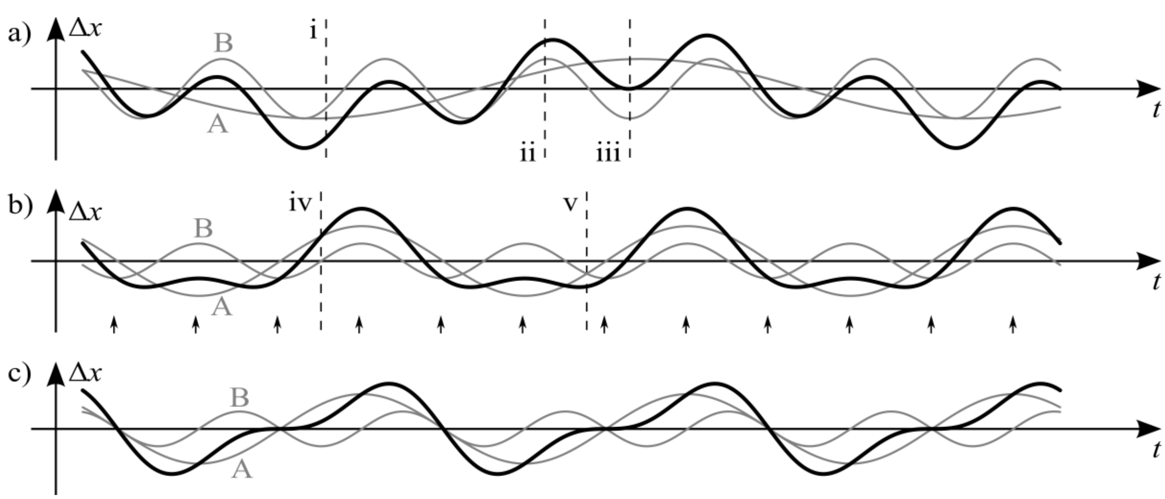


Figure 37.1

Examples of superposition of vibrations. Gray are contributing vibrations, each as it would be without the other present. Thicker black are combined vibrations, the motion that results when both contributing vibrations are present.

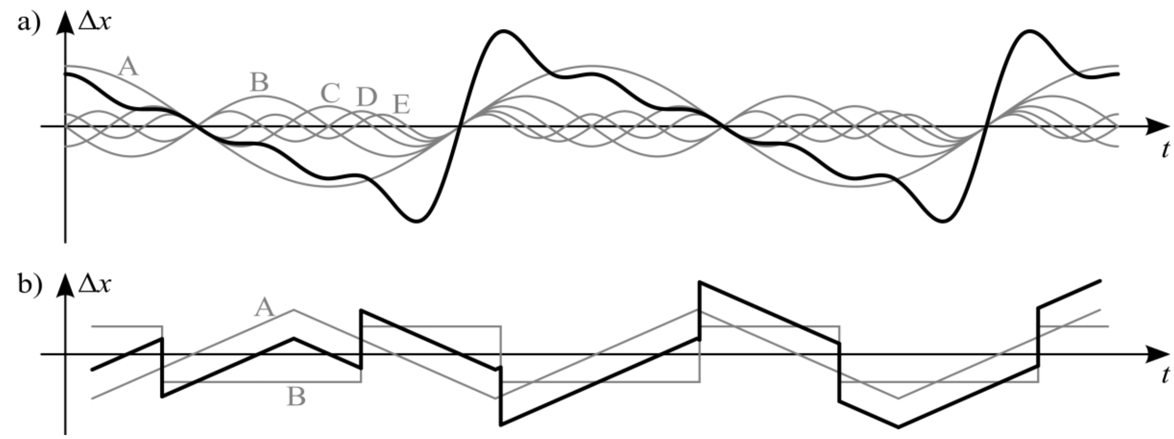


Figure 37.2
More superposition of vibrations. (a) Superposition of five sinusoids. (b) Superposition of two non-sinusoidal vibrations.

Keep in mind that since displacements are measured from the equilibrium position, they can be negative; that must be included when adding. In Figure 37.1(a), at time (ii) both contributing displacements are positive, and at time (i) both are negative; in either case, adding them leads to a displacement of larger magnitude. At time (iii), on the other hand, the contributions are of opposite sign, so that the sum is nearly zero.

Figure 37.1(b) shows the superposition of two pure tones of more similar frequency, and also differing in amplitude. Looking at the combined vibration, it is less obvious that the contributing vibrations were sinusoidal. Part (c) shows the same combination with just one difference: the initial phase of the shorter-period contribution. This change in the phase relationship makes a very significant change in the shape of the combination.

In order to draw a superposition by hand, you must at least find the sum at each maximum and each minimum of all contributing vibrations. For example, in Figure 37.1(b) that would be the 12 times marked by a small arrow at the bottom. In part (c) 18 points would be necessary. (The sinusoids are the same, but they don't line up as nicely.) Once these points are determined, smooth connections between them will at least have the correct general shape. Of course, if the combination has a repeating pattern, as in Figure 37.1(b) and (c), that can cut down the required effort considerably.

At some points in time, it is especially easy to do the sum. When one of the contributing curves is zero, such as time (iv) in Figure 37.1(b), the superposition curve crosses the non-zero contributing curve. Notice that when the contributing curves cross, as at time (v), the combination does *not* go through the intersection.

The examples in Figure 37.1 are a simple type of superposition, combining just two sinusoidal curves. But any number of curves can be combined, and the curves need not be sinusoidal. Figure 37.2(a) shows a superposition of five sinusoids, to make a combined vibration that is pretty far from sinusoidal in shape. Figure 37.2(b) shows the superposition of a triangle vibration and a square vibration with no particular relationship between their periods.

Chapter 38. Superposition with Phase

When looking at the superposition of two sinusoids, how they combine at each point in time is related to the difference in their phases at that time. Figure 38.1 shows two of the superpositions from Figure 37.1, with labels added for the reduced phase of each contributing vibration. Phases for oscillations A (the solid gray lines) are above the graphs and phases for oscillations B (the dotted lines) are below.

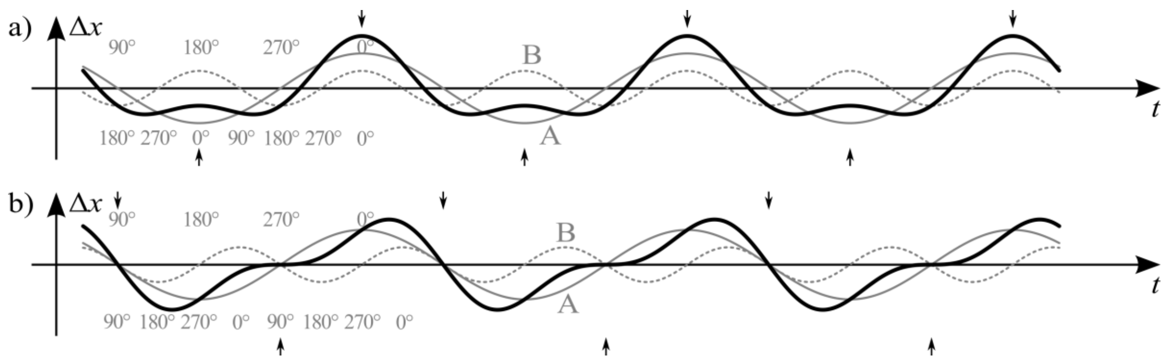


Figure 38.1

Superpositions of sinusoidal vibrations, highlighting phases and phase differences.

Each time when their reduced phases match is marked with a small arrow above the graph. At those points, the (reduced) phase difference between the contributing vibrations is zero,

$$\Delta\phi = \phi_1 - \phi_2 = 0° \quad , \tag{38.1}$$

and the two contributing oscillations are said to be **in phase.** When they are in phase, the two sinusoids are "doing the same thing," and that thing is emphasized in the superposition. In graph (a), at the small arrows from above, they are both at phase 0°, reaching a maximum, so the superposition displacement is especially large. In graph (b), at the small arrows from above, they are both at phase 90°, decreasing steeply, so that the superposition is decreasing with an especially large negative slope. The large displacement may be easier to notice than the large slope, but both are examples of the reinforcement that occurs when vibrations are in phase.

Conversely, the arrows below the graphs in Figure 38.1 indicate where the (reduced) phase difference is 180°, and the vibrations are said to be **out of phase**. At those times the two contributing vibrations are "doing opposite things," and tend to cancel each other. In graph (a), at the small arrows from below, one is at a minimum (i.e., negative displacement) while the other is at a maximum, so that the superposition displacement is relatively small (although not zero, because the amplitudes differ). In graph (b), at the small arrows from below, the contributing slopes are opposed, resulting in zero slope for the superposition.

In the examples of Figure 38.1, the **in phase** and **out of phase** conditions occur only at very specific moments. Although the observations above are true, they don't really make it much easier to understand how the vibrations are superposing. Fundamentally, this is because the phase difference $\Delta\phi$ changes too rapidly in time. In coming chapters, we'll see that when two contributing vibrations have similar periods, then the phase difference changes more slowly, and **in phase** and **out of phase** become more useful ideas.

Chapter 39. Interference

Something a little surprising happens when two sinusoidal vibrations with exactly the same period are superposed. As illustrated in Figure 39.1, the combined vibration is always another sinusoid, with the same period. This will always happen, no matter what the amplitudes and initial phases of the contributing vibrations are. Because of this fact, the superposition of vibrations (and waves) with the same frequency and period is given a special name: **interference.**

Figure 39.1(a) labels the reduced phases at several times for each of the contributing vibrations. An exercise for the reader: which vibration matches the phases at the top, and which matches the phases at the bottom? Since both contributing vibrations go through 360° of phase in the same time interval (the period), their phase difference is the same at any time you choose. For both cases in Figure 39.1 that phase difference is $\phi_A - \phi_B = 90°$.

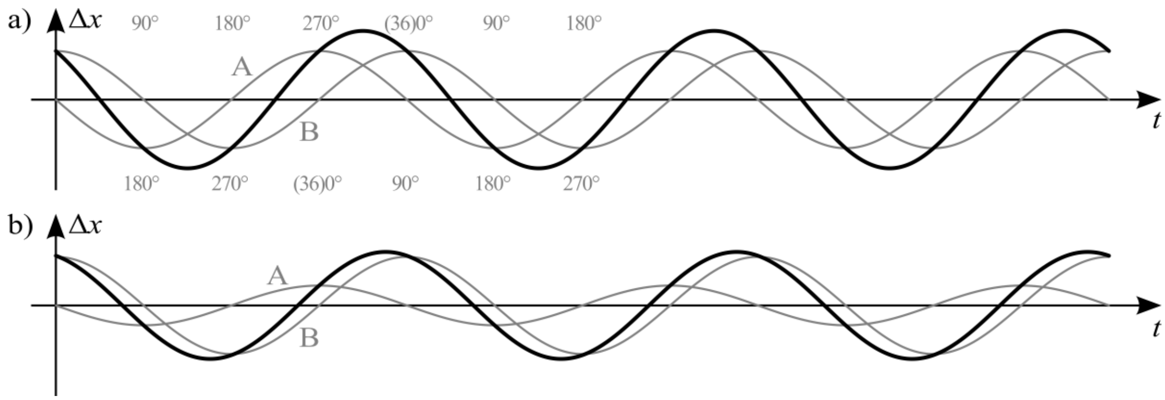


Figure 39.1

Superposition of two sinusoids (gray) with the same period. Result is always another sinusoid (black). (a) Some phases of contributing vibrations labeled; (36)0° represents that 360° is the same as 0°.

This means that, with interference, entire vibrations can be in phase or out of phase, instead of just moments in time. In fact, the name interference presumably comes from the fact that if the contributing vibrations are out of phase, then the superposition amplitude is smaller than at least one of the contributing amplitudes. The two vibrations are "interfering" with one another. Out of phase is the phase relationship that yields the smallest possible superposition amplitude, which is given by

$$x_{\mathrm{m,sup}} = |x_{m1} - x_{m2}| \quad . \tag{39.1}$$

This situation is called **destructive interference.**

But superposition of equal-period oscillations is still called interference even when the oscillations are in phase. In this case the superposition amplitude is the largest possible and is given by

$$x_{\mathrm{m,sup}} = x_{m1} + x_{m2} \quad . \tag{39.2}$$

This situation is called **constructive interference.** So it is called interference even when the two vibrations are reinforcing one another.

Both cases in Figure 39.1 have superposition amplitudes that are larger than either of the contributing amplitudes. It turns out that as long as the phase difference is between $-90°$ and $90°$ (that is, the oscillations are within a quarter cycle of being in phase), then the superposition amplitude will be larger than either amplitude of the contributing vibrations. Conversely, a phase difference between 120° and 240° (that is, within 1/6 cycle of being out of phase) ensures partial cancelation, with the superposition amplitude smaller than the larger of the two contributing vibrations.

Chapter 40. Interference and Energy

Chapter 37 introduces the idea of superposition of vibrations through the question of how a non-resonant sympathetic vibrator would respond to two simultaneous sounds. But the sounds themselves are made of vibrations in the air (Part G goes into detail about that), and the air vibrations superpose before they even reach the resonator. For a specific example, you can think of the air just outside your ear, where the multiple sounds combine before entering the ear canal.

Suppose that you are listening to a pure tone. Let's arbitrarily say that the air just in front of your ear is oscillating with an amplitude $x_{m1} = 2.0\,\mu\mathrm{m}$, and that a cubic centimeter of air there contains $E_1 = 6.0 \times 10^{-14}$ J of sound energy. (Yes, that is a miniscule amount of energy, but it turns out that is a moderately loud sound.) Then suppose a second pure tone is added from a different source, but at the same frequency, so that the superposition is interference. This second tone, if heard by itself, would sound exactly like the first, meaning that it would reach your ear with the same vibration amplitude $x_{m2} = 2.0\ \mu$m and the same energy content. And finally, suppose that they reach your ear in phase, so that constructive interference occurs.

Eq. 39.2 tells us that when combined, the total sound amplitude would be

$$x_{\mathrm{m,sup}} = x_{m1} + x_{m2} = 4.0\ \mu\mathrm{m} \quad . \tag{40.1}$$

According to the general rule about energy and amplitude in proportion 32.1, this means that when both tones are sounding, the cubic centimeter in front of your ear contains an energy E_{sup} given by

$$\frac{E_{\mathrm{sup}}}{E_1} = \left(\frac{\Delta x_{\mathrm{m,sup}}}{\Delta x_1}\right)^2 = \left(\frac{4.0\ \mu\mathrm{m}}{2.0\ \mu\mathrm{m}}\right)^2 = 4 \quad , \tag{40.2}$$

$$E_{\mathrm{sup}} = 4E_1 = 24.0 \times 10^{-14}\ \mathrm{J} \tag{40.3}$$

How can this be? We appear to have added together two things with a certain amount of energy and come out with four times as much energy in total, double the simple sum of the energy. Doesn't this violate the idea of energy conservation, as described in Chapter 29?

The resolution of this paradox involves phase differences. Somewhere else, a moderate distance away from your ear, there is some place where the two pure tones are combining out of phase. In that location, the amplitudes are combining according to Eq. 39.1 as

$$x_{\mathrm{m,sup}} = x_{m1} - x_{m2} = 0 \quad , \tag{40.4}$$

so that in that place the energy is mysteriously disappearing, instead of mysteriously overabundant.

When all the different places where the sounds are overlapping are taken into account, it works out that energy is indeed conserved. The places with an excess of energy are balanced by those places with a lack of energy. In this chapter, we will not learn about why the phase difference varies in different locations. (See Chapter 156 for more about that.) The point here is simply that in order for amplitude superposition and energy conservation to coexist as correct physical models, some pretty subtle things must happen.

Practically speaking, whenever we are faced with a question about combining sounds "locally" (i.e., at a point or in a small region), amplitude superposition will be the most useful model. Also, when a variety of frequencies are involved, energy conservation will work out "on average." The trickiest situations are interference (combining several sounds with the same frequency), in which cases energy conservation can lead you astray if you aren't very careful.

Chapter 41. Beats

41a. Two Equal Amplitudes

When you listen to two sounds that have slightly different periods, an interesting and useful effect occurs. Figure 41.1 illustrates the effect when two pure tones have the same amplitude. The two contributing sinusoidal vibrations are in part (a), with periods related by

$$T_2 = \frac{20}{21} T_1 \quad , \tag{41.1}$$

where subscripts 1 and 2 refer to the solid and dotted curves respectively. They start in phase, but after 10 cycles of the solid curve they are out of phase. The result is that in the superposition, shown in part (b), the amplitude (indicated by the dashed line) gets alternately larger and smaller. Since amplitude correlates to loudness, the effect is clearly audible as a fluctuating loudness called **beats.**

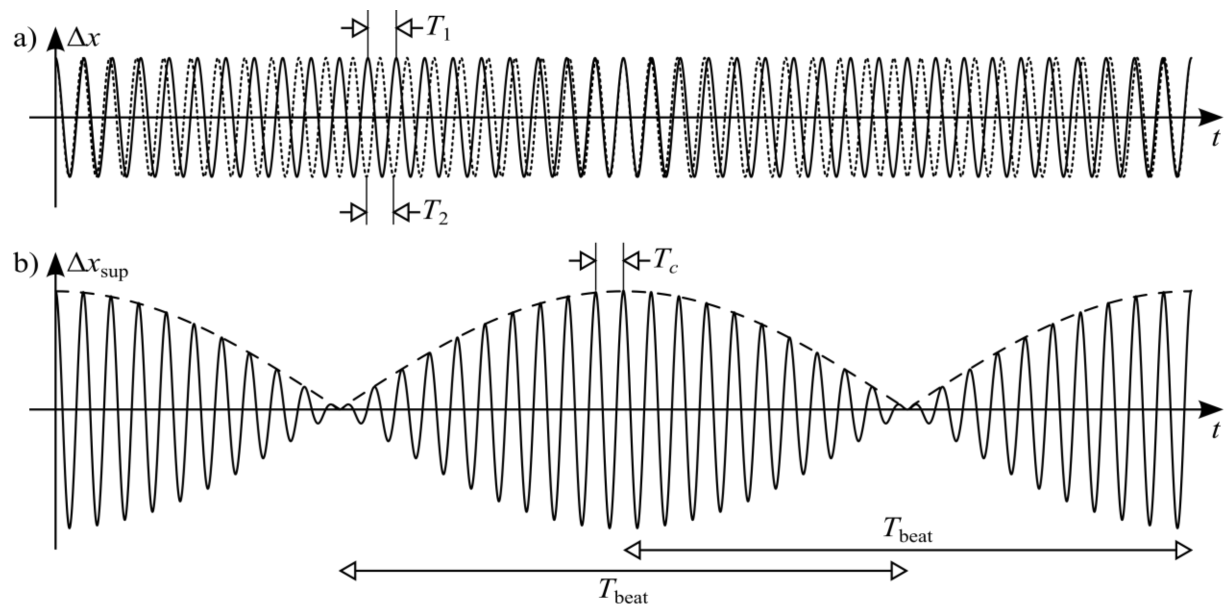


Figure 41.1

Superposition with very similar periods, resulting in beats. (a) Individual vibrations. (b) Their superposition.

Some auditory effects have their origin in the physiology of the ear and brain, but beats are pure physics. It can be shown that the superposition is exactly a sinusoidal vibration with variable amplitude. Showing this mathematically requires trigonometry, beyond the scope of this book. But the results of the mathematical derivation are easy to understand. The frequency of the combined vibration, which determines the perceived pitch, is the average of the two contributing frequencies,

$$\frac{1}{T_c} = f_c = \frac{1}{2}(f_1 + f_2) \quad , \tag{41.2}$$

where the subscript was chosen to be short for "combined." The amplitude varies periodically, as the absolute value of a sinusoid. The time between moments of maximum loudness (or between minimum loudness) is the period of one **beat**. The same trigonometry shows that the associated beat frequency is simply the difference between the two contributing frequencies,

$$\frac{1}{T_{\text{beat}}} = f_{\text{beat}} = |f_1 - f_2| \quad , \tag{41.3}$$

where the absolute value has been added in case $f_2 > f_1$.

Humans can hear the beating as a loudness variation if the beat frequency is below roughly 15 Hz. If the beat frequency is higher than that, then the same math still applies, but the loudness variation is too rapid to be perceived.

41b. Extra: Unequal Amplitudes

A different way for a sound to deviate from a pure tone is for the pitch to vary slightly. When the frequency of a sound shifts up and down periodically, the variation is called a **warble.** This is different from beats: warble means varying frequency, beats mean varying amplitude.

If the two pure tones in Figure 41.1(a) had different amplitudes (implying that they would have different loudness if heard individually), then much the same thing would happen as with equal amplitudes. The pure tones would still shift in and out of phase, resulting in a larger and smaller combined amplitudes respectively. The smallest amplitude would not be zero, but the combined loudness would still vary, and those beats would occur with exactly the frequency given by Eq. 41.3.

In the case of unequal amplitudes, there is *also* a slight warble in the combined sound. To show this mathematically requires advanced trigonometry, and in fact it is rarely discussed even in physics courses. The warble frequency, that is, how often the frequency variation repeats its pattern, is the same as the beat frequency. The details of how the frequency varies are complicated. For most of the cycle, the combined frequency is between the average frequency (as from Eq. 41.2) and the frequency of the larger amplitude pure tone.

However, this particular type of warble is generally difficult to perceive, for various reasons. For example, suppose that you can make two pure tones, one with a frequency of 500 Hz, and another with a frequency of 504 Hz and one third the amplitude of the first. Those frequencies are so close together that it is very difficult for the human ear to tell them apart (see Chapter 81 for more details). If those two tones were sounded at the same time, the combination would beat and warble at a rate of $f_{\text{beat}} = 504\text{ Hz} - 500\text{ Hz} = 4\text{ Hz}$. But the warble would mostly be between 501 Hz and 500 Hz (with periodic dips to 498 Hz), a variation in pitch that is just too small to hear. The beats, on the other hand, would be clearly audible, with the largest amplitude twice as large as the smallest amplitude (see Eqs. 39.1 and 39.2).

If the two pure tones were farther apart, say 500 Hz and 520 Hz, then the warble would be larger too, varying from 505 Hz to 500 Hz with periodic dips to 490 Hz. That variation in pitch is large enough to perceive. But because it would vary at a rate of $f_{\text{beat}} = 520\text{ Hz} - 500\text{ Hz} = 20\text{ Hz}$, which is greater than 15 Hz, it would happen too rapidly to perceive. The beats would be too rapid to perceive as well. The combined sound would just have a buzzing quality.

In summary, although superposition of similar frequency tones with different amplitudes causes a warble, that warble is nearly always imperceptible. Conversely, if you hear a sound with a warble, it probably arises from some other reason than superposition. On the other hand, beats (meaning a periodic variation in loudness) can be quite perceptible, regardless of whether the combining amplitudes are equal or not.

Chapter 42. Fourier Analysis

The vibrations associated with audible sounds have periods so short that it is impossible to directly observe their motion. But they can be observed indirectly, using a microphone to convert the motion into an electrical signal and an oscilloscope to observe that signal. With some experimentation it is soon seen that among complex periodic vibrations, just as with SHM, the period and frequency still primarily determine the perceived pitch, and the size of a typical displacement primarily affects the loudness. (Complex vibrations, even periodic ones, can easily be so complicated that precisely what the term *amplitude* should mean is not clear.)

Different timbres are seen to result from different shapes of each cycle of the displacement versus time graph. Generally, vibrations with more wiggles in each cycle produce harsher sounds, compared to the pure tone produced by SHM. Also, displacement graphs that have sharper corners produce harsher sounds. These qualitative observations are fine as far as they go, but how are we to numerically and completely characterize all possible vibrations?

In the 1820s, French mathematician Joseph Fourier provided a good solution with the following proposal, which was later rigorously proven and is now known as **Fourier's theorem**.[7] It is the primary reason that sinusoidal motion is so important to acoustics.

[7] Joseph Fourier, *Théorie analytique de la chaleur* (Paris: Firmin Didot Père et Fils, 1822).

> Any physically possible complex vibrational motion can be built up by superposition from sinusoidal vibrations. You never need more than one sinusoidal vibration at each frequency.

So, to understand the timbre of a vibration, we can do the superposition process in reverse: given a complex vibration, break it down into a set of sinusoidal vibrations that would superpose to make it. This process is **Fourier analysis**, and the mathematical operation that accomplishes it is called a **Fourier transform**. The math underlying the Fourier transform is beyond the scope of this book, as it requires calculus. In practice, one usually knows the shape of a complex vibration from measuring the displacement at a finite number of points in time. In that case, there is a particularly efficient algorithm called the **fast Fourier transform**, or **FFT**. There are many software programs and scientific instruments that can take FFTs of sound measurements.

To completely display the results of Fourier analysis, we only need to show, for each of the contributing SHMs, the three parameters that describe it: frequency f_n, amplitude x_{mn}, and initial phase ϕ_{0n}. The subscript *n* here is a placeholder for an index to identify the different contributing SHMs. Notice that Fourier analysis neatly sidesteps the problem that amplitude is not clearly defined for a complex vibration. Since we break down a complex vibration into a set of SHMs, we only need to define amplitude for SHM.

If you have read Chapter 39, you have seen why there never needs to be two contributing SHMs with duplicate frequencies; otherwise, you can just take Fourier's word for it. In any case, this fact suggests graphing the other parameters along a frequency axis. But graphs only have one vertical axis, and we have two remaining parameters. This problem is solved by Ohm's acoustic law. (Georg Ohm, a German physicist, proposed this in the 1840s.[8] You may be familiar with his more famous electrical law.)

> The sound perceived by human ears depends only on the frequencies and strengths of the constituent sinusoidal vibrations. The initial phase does not affect what we hear.

This law is not the least bit obvious. First, although initial phase of one SHM is not physically meaningful and cannot be perceived, in superposition the relationships *between* the initial phases *do* have an effect on the vibration shape. In fact, with many contributing SHMs, varying their initial phases can change the vibration shape quite radically. In principle, it could be possible for our auditory system to detect such changes.

And yet, for the most part, we don't. Very different vibrations that differ only in the initial phases of their underlying SHMs all sound pretty much the same to us. Ohms' Acoustical Law is not perfectly accurate; such shape changes can lead to subtle perceptible differences. But for the most part, initial phase is not important in determining timbre.

Being able to ignore initial phase leaves the vertical axis free for amplitude. To describe how a complex vibration sounds, we just need to show the frequency and amplitude of each of the contributing SHMs. This could look something like Figure 42.1(a), in which each of four SHMs are represented by a point. A graph of

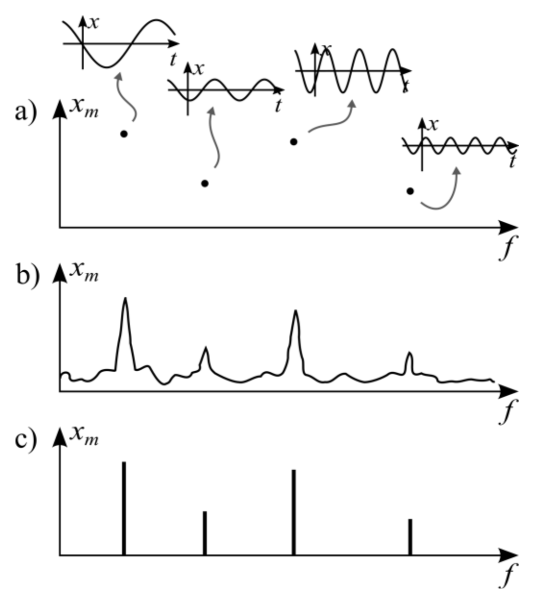


Figure 42.1

Different appearances for an amplitude spectrum of complex vibration. (a) Each point/peak represents a sinusoidal time curve.

[8] Georg Ohm, "," *Annalen der Physik* 59 (1843): 519.

anything versus frequency is called a **spectrum.** The graphs in Figure 42.1 might be called **Fourier spectra** to indicate that they are the result of Fourier analysis, or **amplitude spectra** to describe what is on the vertical axis.

In a laboratory situation measuring an unknown vibration, with unknown frequency components and some imprecision of measurement, an FFT is more likely to produce a graph like Figure 42.1(b). That curve has a value at *every* frequency, although the same four points from part (a) are evident as peaks. Many real sounds have spectra like this, with clear peaks at specific frequencies. So, in this book they will be represented as in Figure 42.1(c). The features in the graph will be called **peaks,** even though they are drawn as vertical lines. This is somewhat like representing a person with a stick figure.

Notice that in all of these spectra, only the positive half of each axis is shown. Neither negative frequencies nor negative amplitudes make any sense.

In summary, different sound timbres can be understood as arising from different shapes of displacement versus time, but they can also be understood as arising from different sound spectra, and in particular from different patterns of peaks in the spectra. Focusing on spectra simplifies things in two ways. First, spectra do away with the initial phase information, which is extraneous according to Ohm's Acoustical Law. Second, spectra often can be described much more simply than a displacement graph, because they often consist of a few discrete peaks.

These two different ways of looking at vibrations are called the **time domain** and the **frequency domain.** Neither is more correct than the other. But some effects and consequences are more easily understood in the time domain, while others are more easily understood in the frequency domain. It is very useful to an acoustician to be able to adjust their thinking from one domain to the other with ease.

Chapter 43. Spectrum Parts

Many sounds that one encounters have spectra that consist of well-defined peaks. These are sounds that have some tonal quality to them, although they may not have a clear pitch, as with gongs. Each peak is termed a **partial** of the spectrum. In order to refer to a specific partial, they are numbered from lowest frequency to higher frequencies, starting with 1 as in Figure 43.1.

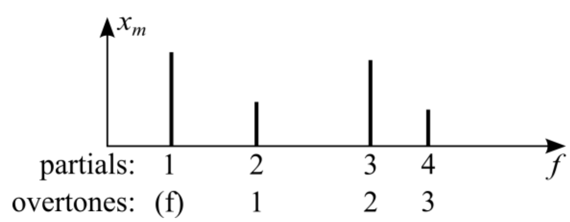


Figure 43.1

Spectrum of generic vibration with peaks, numbered as partials and overtones. (f) indicates the fundamental.

The lowest-frequency partial is called the **fundamental,** especially in the common case that this peak has one of the largest amplitudes. The fundamental often plays a primary role in determining the perceived pitch of the sound. All the other peaks are called **overtones.** When referring to the overtones specifically, they are also numbered from 1 at the lowest frequency towards higher frequencies (see Figure 43.1). Thus, the first overtone is the second partial, the second overtone is the third partial, and so on.

Even though each peak (or partial) in a spectrum represents a sinusoidal vibration, which has a definite period, a spectrum consisting of several peaks might describe a complex vibration that is *not* periodic. Unless two partials have periods with a special relationship, each time that the first partial sinusoid reaches maximum displacement, at reduced phase $\phi = 0$, the other partial will be at some new and different phase, with the result that the superposition never repeats. Or, the combination may repeat only after a very long time, which has the same implications for perceived sound. The special relationship needed to keep the complex vibration periodic is explored in Section 44a.

The partials are said to combine to make the full complex vibration. But notice that this does not mean that any characteristics of the peaks combine mathematically in any way. There is no meaningful sense in which their frequencies could be added, for instance. The only mathematical combination relating to a spectrum is the implied sum of displacements in the superposition, as described in Chapter 37. However, when working with spectra, one cannot actually perform that addition, because the phase relationship information is missing. The amplitudes of the partials cannot even be summed meaningfully; that *might* yield the maximum displacement of the complex vibration, but only if the SHMs of the partials all happen to reach phase 0° at the same instant in time, and there is no particular reason to expect that to happen.

Chapter 44. Special Spectra

44a. Harmonic Spectra

If a complex vibration is periodic, then it is called a harmonic sound, perhaps because this sound type conveys the strongest sense of pitch. These are the sorts of sounds that are primarily found in music, excepting percussion. The periodicity of the complex vibration has significant consequences for the frequencies of the partials in its spectrum.

If the complex period is T, then every partial SHM must also repeat each time interval $\Delta t = T$. This is due to a property of sinusoids: if any one partial SHM were to disobey the T periodicity, then it isn't possible for sinusoids of other frequencies to repair the damage.

This doesn't mean that T is the only period allowed, however. Each partial is required to have a whole number of cycles in time T, but more than one cycle is still an option as shown in Figure 44.1. So, the allowed periods follow the nice pattern $T_1 = T$, $T_2 = T/2$, $T_3 = T/3$, … Even nicer is the pattern of frequencies, $f_1 = f$, $f_2 = 2f$, $f_3 = 3f$, … That pattern can be summarized by the single equation

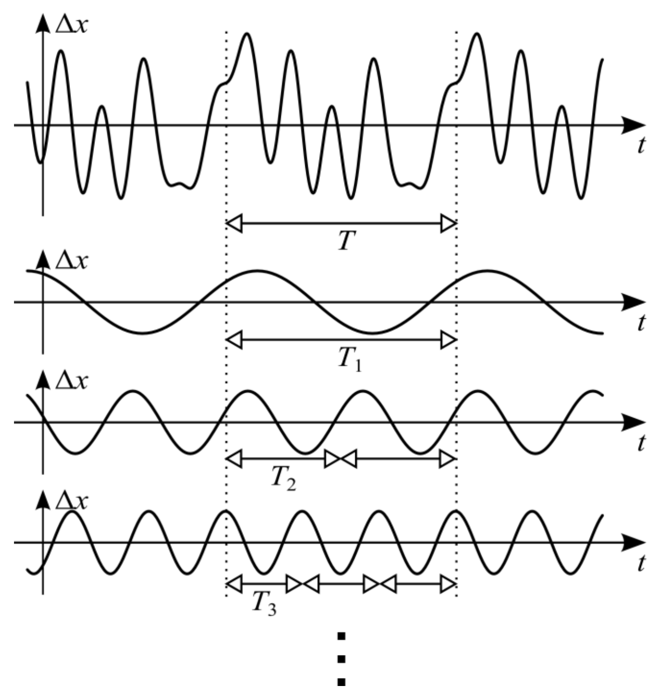


Figure 44.1
Periodic complex vibration and some of its partials, showing relationship among periods.

$$f_n = nf \quad , \tag{44.1}$$

where f is the frequency of the complex vibration, f_n is the frequency of one partial in the spectrum, and n is the **harmonic number** of that partial, which must be a positive whole number.

The partials of a harmonic (that is, periodic) vibration, which follow this pattern, are called **harmonics**. They are typically identified by their harmonic number, so that the partial with frequency f_n in Eq. 44.1 is called the n^{th} harmonic. The fundamental frequency, which is the same as the frequency of the original complex vibration, is the first harmonic.

If you have read Chapter 179, you may recognize Eq. 44.1 as essentially identical to Eq. 179.2. The relationships and jargon here are just the same as for the modes of 1D standing waves. Partial number here plays a role very similar to mode number in Chapter 179.

Harmonic spectra are often readily recognizable by their equally spaced peaks. However, it is not required that there be partials at every harmonic frequency. Or, to say the same thing in a different way, some of the harmonics may have zero amplitude. Figure 44.2 shows the amplitude spectrum for the vibration in Figure 44.1, which has no partial at frequency $4f$. Notice that as a result, for higher frequencies the partial number does not match the harmonic number.

Figure 44.2

Spectrum of the harmonic vibration from Figure 44.1, with peaks labeled as partials and harmonics.

It is even possible for a harmonic spectrum to have no peak at the frequency that all others are multiples of, as shown in Figure 44.3. Now we have a conflict about which frequency to call "fundamental." Should the fundamental be the peak in the spectrum with lowest frequency, or should the fundamental be the frequency that all the others are multiples of? The usual resolution in this case is to say that the spectrum has a **missing fundamental.** The frequency labeled f_1 in Figure 44.3 is called the fundamental, and the other frequencies continue to be multiples of the fundamental frequency. However, the numbering of partials only includes actual peaks, so that the first partial is no longer the fundamental.

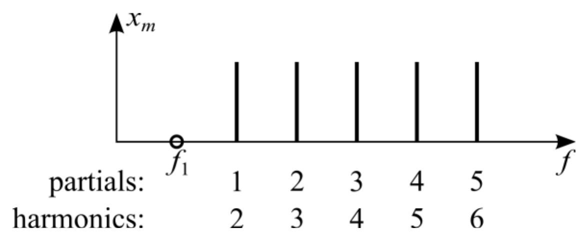


Figure 44.3

Spectrum of harmonic vibration with missing fundamental.

The simplest of all spectra consists of a single peak, the spectrum of an SHM or a pure tone. Pure tones and harmonic sounds both cause a strong sense of pitch. A pure tone is not necessarily a member of the harmonic sound category: it can't have equally spaced peaks, since it doesn't even have multiple peaks. But a pure tone could be considered to be a harmonic sound with zero amplitude for all overtones, sort of the exact opposite of Figure 44.3.

As a final note, some sounds have partial frequencies that are close to whole-number multiples of a fundamental, but not quite exact multiples. The sound from piano strings is one example. In these cases peaks are often still called harmonics. There is no precise rule about how far a spectrum can stray and still be called harmonic. If you are in doubt, stay on the safe side and describe the peaks as "partials" instead of "harmonics."

44b. Extra: The Square Wave

A specific example of a harmonic spectrum with missing harmonics deserves particular mention. The so-called **square wave** is depicted in Figure 44.4. Despite the traditional name, this is not truly a wave, as defined in Chapter 112; it is an oscillation. An ideal square wave vibration alternates between two displacements, switching instantaneously between them. This is physically impossible, but there are some situations where the transitions are very fast, so that the square wave is a good model.

Due to its vertical segments and sharp corners, it takes the superposition of an infinite number of sinusoidal partials to make the square wave. But the higher frequency partials have decreasingly small amplitudes, so that a superposition of the lower frequency partials can approximate the square wave pretty well. Figure 44.4(a) shows an ideal square wave in black, and the superposition of the first five partials in gray. It could even be argued that the finite-partials approximation is a better model for real oscillations, since the transitions are not instantaneous.

Figure 44.4(b) shows the spectrum of a square wave, which evidently consists of only odd harmonics, skipping all the even-multiple frequencies. This can be deceptive, because the equally spaced peaks can be

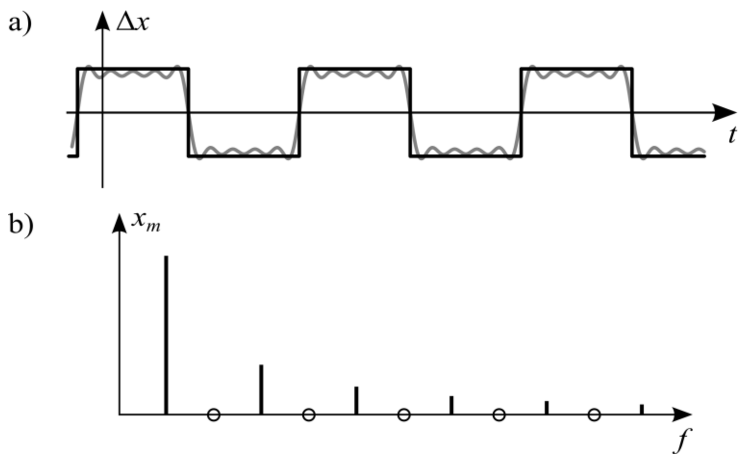


Figure 44.4
(a) Ideal square wave and approximation built of the first five partials.
(b) Square wave's amplitude spectrum, circles marking missing harmonics.

mistaken for a normal harmonic spectrum. In order to properly recognize a harmonic spectrum, it is important to spot not only equally spaced peaks, but also a compatible spacing of the fundamental from the origin or vertical axis.

The square wave spectrum also illustrates a common feature of many real spectra, that the partial amplitudes decrease as the frequency increases. The particular rule for the square wave is

$$\frac{x_{mn}}{x_{m1}} = \frac{1}{n} \quad . \tag{44.2}$$

Keep in mind that, as interesting as the patterns in a square wave are, this is a very special case. Most harmonic spectra do not skip so many harmonic frequencies. When working a question with a generic harmonic spectrum, the best assumption (barring additional information) is that all harmonics are present in the spectrum.

If you have read Chapter 181, then you may recall another situation calling for only odd harmonics: the standing wave modes of a 1D medium with fixed and free ends. This is ***not*** to say that a square wave is somehow made of standing waves. The situations are quite different; a "square wave" isn't even really a wave. This point of similarity, both things involving odd multiples, is more of a coincidence, perhaps useful for memorization.

44c. Extra: Time Average Spectra

If the same source can produce several sounds at different pitches, they are usually recognizable as having the same timbre. How do we group the sounds from one musical instrument as having the same timbre, distinct from another instrument? How do we recognize different people based on their voices? More generally, what does it mean to change pitch while keeping timbre the same?

Chapter 42 describes how timbre is related to the "shape" of a sound. Sounds with different timbres have differently shaped displacement versus time graphs, and also differently shaped spectra. However, as described by Ohm's Law, not all differences in displacement graph shape result in a timbre change. Figure 44.5(a) shows two periodic sounds (in gray and in black) with quite different displacement graphs, but the same spectra. The sounds differ only in the initial phases of the component SHMs, and Ohm's Law tells us that they will sound the same to the human ear. This suggests that the shape of the spectrum is more important than the shape of the displacement graph.

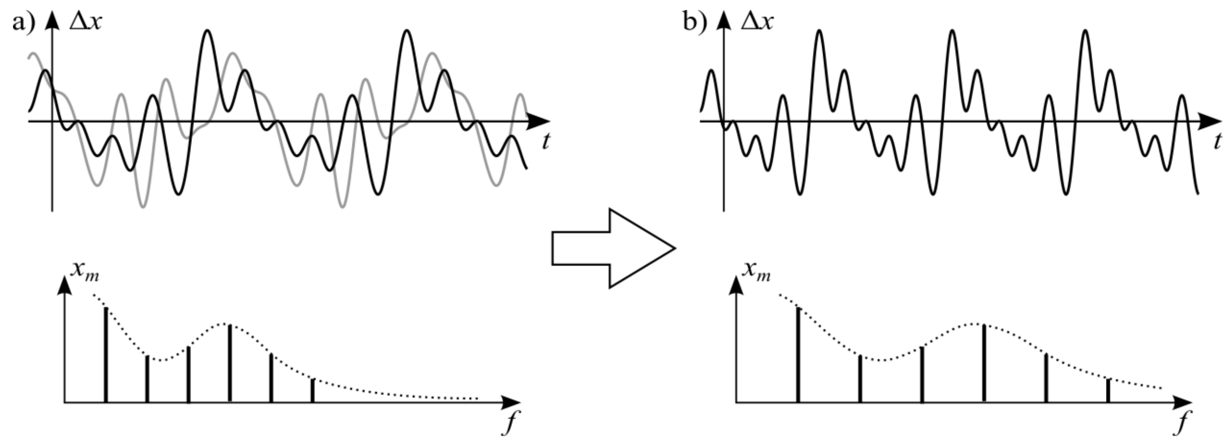


Figure 44.5

Changing pitch of a sound by scaling time characteristics. (a) Gray and black traces have same spectrum, partials differing only in phase constant. Period in (b) is 2/3 of period in (a).

If maintaining the same timbre requires keeping the same overall shape of the spectrum describing the sound, one obvious possibility presents itself. Figure 44.5 illustrates changing a sound by simply speeding it up, from the black displacement graph in (a) to the curve in (b). All time measurements are reduced by some factor (2/3 in the figure), including the period of the complex oscillation and the periods of all the partials. The effect on the spectrum of the oscillation is to increase all frequencies by the inverse of that factor (3/2 in the figure).

In the amplitude spectra of Figure 44.5, the partials are shown along with a dotted line **envelope** of the spectrum. An envelope is not itself part of a graph; rather, it describes a region of a graph that the actual spectrum stays within (just as a written letter stays inside the envelope used to mail it). The envelope helps us see the shape of the spectrum, even while the spectrum itself is just peaks at specific frequencies. The increase in frequencies is evident as a horizontal stretching of both the spectrum and envelope. For both the oscillation graph and the spectrum, the change shown qualifies in one sense as "keeping the same shape."

If you have ever had the opportunity to play with a record player, this is exactly the sort of effect that is achieved by playing a vinyl record at the wrong speed (e.g., 45 rpm instead of 33 rpm). And if you have had that opportunity, you also know that this does *not* keep the timbre the same. When this is done to the human voice, it creates a sound with the timbre made famous by *Alvin and the Chipmunks.* Evidently, this way of "keeping the same shape" is not what we need to keep the same timbre.

Figure 44.6 shows a different way to increase the pitch of a sound. The key to keeping the same timbre is that the spectrum envelope stays the same. In the spectra, the frequencies of the partials are being changed in the same way as in Figure 44.5: the fundamental has increased (by a factor of 3/2), and the overtones remain harmonically related. (Notice that remaining harmonic does *not* mean that the spacing between the partials stays the same.) However, the amplitudes of the partials have also changed, in such a way that they stay within the same envelope.

In the displacement graph, the shape of the oscillation has not stayed the same. The period is 2/3 shorter, as it must be for the higher pitch, but in addition each cycle has a different shape. But then, as Figure 44.5(a) shows, the displacement graph shape is not a good indicator of whether timbre has stayed the same.

In summary, different sounds have the same timbre if their spectra stay within the same envelope. That's is a rather slippery rule, though. As illustrated in Figure 44.7, a single spectrum is not enough to define a particular envelope. This is part of what makes timbre so difficult to pin down, not only in terms of a quantitative description, but even for our perception of it.

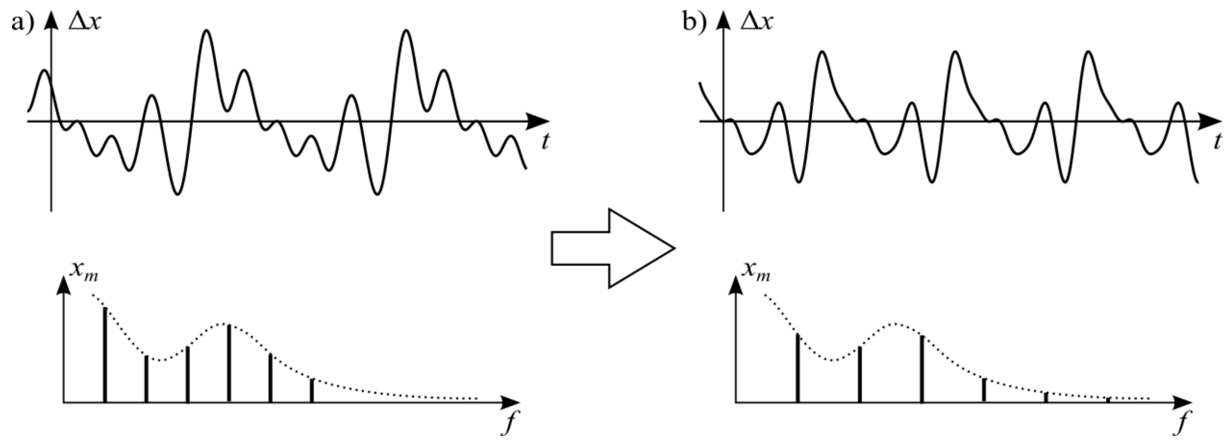


Figure 44.6
Changing pitch of a sound while keeping same timbre.

Timbre is better understood to be a property of a particular source, as opposed a property of a specific sound. In order to measure the timbre of a source, we need to measure the spectra of many different pitches and sounds from the source, which are then combined to determine the envelope within which they all fit. This envelope is called a **time average spectrum.**

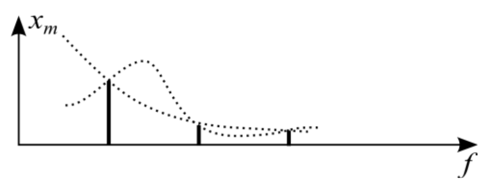


Figure 44.7
Two possible envelopes for the same spectrum.

Given the time average spectrum for a musical source, we can then predict a spectrum for any given pitch. The pitch determines the frequency of the fundamental, the fact that the source is musical specifies that the overtone frequencies are harmonics, and the relative heights of the partials are chosen to match the time average spectrum. The spectra for various loudness levels can then be predicted by multiplying all the partial peak heights by an overall factor, growing or shrinking them while keeping the same height relationships.

This prediction method can never be perfect, however. Sources often have timbres which vary a bit with loudness. Indeed, timbre can often change over time. Even a simple tuning fork sounds different when first struck than it does after it has been ringing for a few seconds. The situation is even further complicated for non-musical sources. Time average spectra are useful to give an overview of the sort of spectra that a source tends to produce, but their use as a predictive tool is only suggestive.

The idea of averaging spectra is sometimes extended to **long-time average spectra (LTSA),** which average sounds that may not even maintain the same timbre. For instance, one might find the LTSA for human speech, or for the background noise and sounds in some particular environment. An LTSA gives an idea of how the studied sound environment is spread across the frequencies, but it is not associated with any particular sound, let alone any particular pitch of any particular sound.

Chapter 45. Noise

There are some sounds that do not have distinct peaks in their spectra. In fact, when measuring spectra in the real world, indistinct spectra are quite likely, because it can be difficult to get samples of single sounds without background. But measurement difficulties aside, there are some sounds that intrinsically do not have sharp peaks in their spectra. This is the technical definition of **noise.**

Without peaks, the "stick figure" spectrum representations aren't appropriate. And it is no longer as useful to think of these spectra as recipes for the superposition of a limited or organized set of sinusoidal motions.

A more useful, if more vague, conception is that these spectra show "how much" of each frequency is present in the sound.[9]

Figure 45.1(a) shows **white noise,** which has equal contributions at all frequencies. The classic example of white noise is the hissing sound heard when a radio is tuned between stations. (Some people may be familiar with this classic example. But some may not be, because many modern radio tuners block out this sound when a station signal is not detected at the tuned frequency.) Because noise is in a sense disorganized, any given measurement will give a ragged spectrum, represented by the gray line in Figure 45.1(a). But on average, the spectrum is the featureless horizontal line.

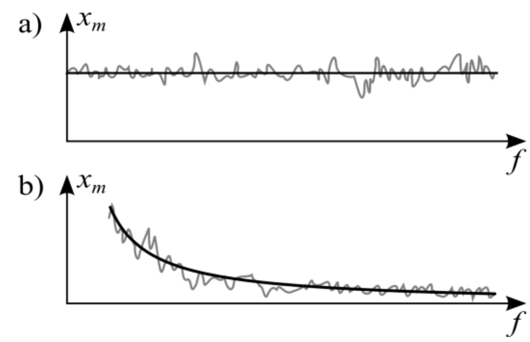


Figure 45.1

Spectra of noise. (a) White noise. (b) Pink noise.

Figure 45.1(b) shows **pink noise,** which is arranged so that it has equal contributions in all musical ranges, such as bass, mid-range, and trebble. This may seem the same as the description of white noise; the difference is due to the way that we perceive pitch, as described in Chapter 77.

When measured in the time domain, noise will show no recognizable pattern. The displacement changes randomly with time, and it may be impossible to even precisely define an amplitude. Types of noise with different spectra may have subtle distinguishing characteristics. For example, white noise, with larger amplitudes at higher frequencies will appear more jagged than pink noise. But overall, noise is easier to characterize in the frequency domain.

Due to the lack of peaks, noise sounds generally do not have a distinct pitch. But they may give an overall sense of lower or higher pitch, depending on which frequency regions are more prominent in their spectra. Other examples of noise include applause, and the "sh" phoneme.

Some extremely brief sounds, such as a "pop" or a drum beat, also do not have a distinct pitch. But these sounds are not quite the same as noise. The problem here is that the duration is too short to even establish a period of the motion at all, as there is no repetition. This book will not attempt to address such sounds.

Chapter 46. Spectrograms

46a. Spectrogram Construction

A spectrum is only able to represent a sound that is uniform and unchanging (or nearly so). A time averaged spectrum (see Section 44c) can describe a changing sound, but only in an overall sense. To graphically represent a sound that changes over time, we can use a **spectrogram**. This chapter gives a very basic introduction to the topic.

The key consideration is that in order to represent change over time, one dimension or axis of the two-dimensional page must be used to represent that time. This leaves only one dimension to represent the sound at each moment, something that in preceding chapters was accomplished with a two-dimensional spectrum graph. The solution in spectrograms is to represent amplitude not with a vertical axis, but with color.

Figure 46.1 shows examples. Spectrum (a) is similar to the pink noise in Chapter 45, and in the bar below the amplitude is represented by darkness, with the largest being black. The noise does not have specific

[9] For the purist: In fact, the vertical axis for such spectra is not truly amplitude. Describing the vertical axis properly is beyond the scope of this book.

peaks, so the color transitions slowly along the frequency axis. Since audio software and other sources of spectrograms need to be able to represent any sort of spectrum, they can show colors that blend like this.

But for the stick-figure spectra that are the focus in this book, each partial can be represented by a colored spot. In Figure 46.1(b), the spots are colored to represent amplitude, just as in part (a).

One-dimensional spectra like these can then be stacked one after the other to illustrate changing sounds. As seen in Figure 46.2, the usual convention is to use a horizontal time axis. Each 1D spectrum is rotated 90° to make the frequency axis vertical.

- The spectrum from Figure 46.1(b) is marked in this spectrogram by the thin rectangle. But we can now see that this same spectrum is heard throughout the first second of time.
- During the next second of time, a different sound occurs. Its three partials have amplitudes similar to the first sound, and the third partial has the same frequency as in the first sound. This sound is harmonic, because the frequencies of the partials are multiples of the fundamental frequency.
- During the last second, the sound is the same as in the middle second, expect for having a lower intensity. The change in color may be difficult to see, but all the lines are less dark.

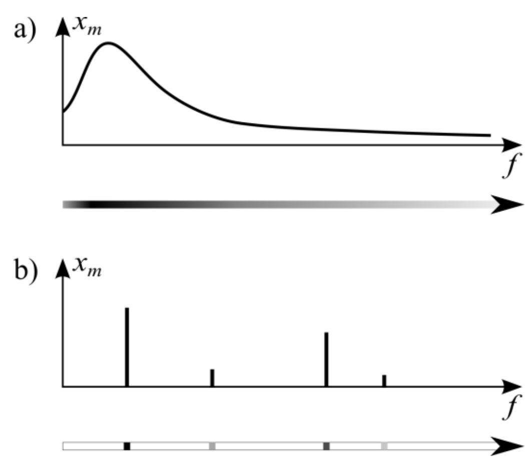


Figure 46.1

Spectra and corresponding one-dimensional representation. (a) Noise. (b) Sound with distinct partials.

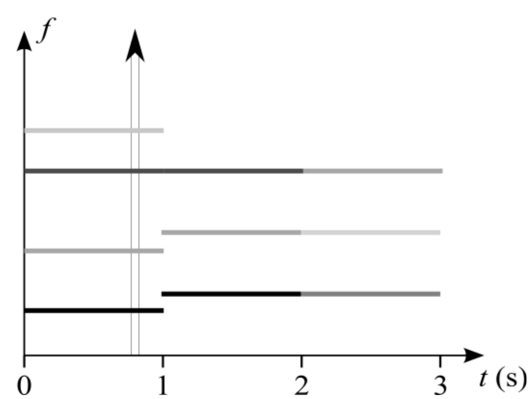


Figure 46.2

Spectrogram of three successive sounds, the first matching the spectrum of Figure 46.1(b).

46b. Extra: Time and Frequency Uncertainty

The two axes of a spectrogram show different types of information, which we perceive as pitch and duration, but both relate to changes over time. This is explicit for the time axis. But frequency can only be determined from the rate of cycle repetition, which also requires time to establish. The variation of a displacement versus time can only hold a limited amount of information, which leads to a fundamental "fuzziness" in spectrograms. This can be surprising, because we perceive pitch and duration so differently.

To illustrate the difficulty, Figure 46.3 shows progressively shorter time intervals of the same vibration. The bold curves of graphs (b–d) are exact duplicates of portions of graph (a), while the lighter curves bring the displacement promptly back to the equilibrium position. In one sense, these all have the same period (and thus frequency), since they are all segments of the same sinusoidal graph. However, if you try to actually measure the period, it would get progressively more difficult for graphs lower in the figure. As a result, you probably would have less and less confidence in your measurement result.

This lack of confidence is not simply due to limitations of humans in interpreting graphs. There is a mathematical sense in which the frequency of Figure 46.3(d) is simply not well defined. The mathematical details require a much more thorough treatment of Fourier analysis than will be given in this book. But the result is that associated with each of these graphs there is both a best choice for the frequency f and a **frequency uncertainty** Δf. The frequency of the graph is between $f - \frac{1}{2}\Delta f$ and $f + \frac{1}{2}\Delta f$, but it is impossible to be more precise than that.

The relationship between frequency uncertainty and the duration of the graph is called the **Gabor limit**,

$$\Delta f \, \Delta t \geq 1 \quad . \qquad (46.1)$$

When the duration is long, including many cycles, then the frequency uncertainty is relatively small. But if the duration is only one period long, then the frequency uncertainty becomes as large as the frequency itself! The Gabor limit is an inequality because some sounds never have a clearly defined frequency, no matter how long you listen. The noise represented in Figure 46.1(a) would never have a frequency uncertainty smaller than the width of the broad peak in the spectrum.

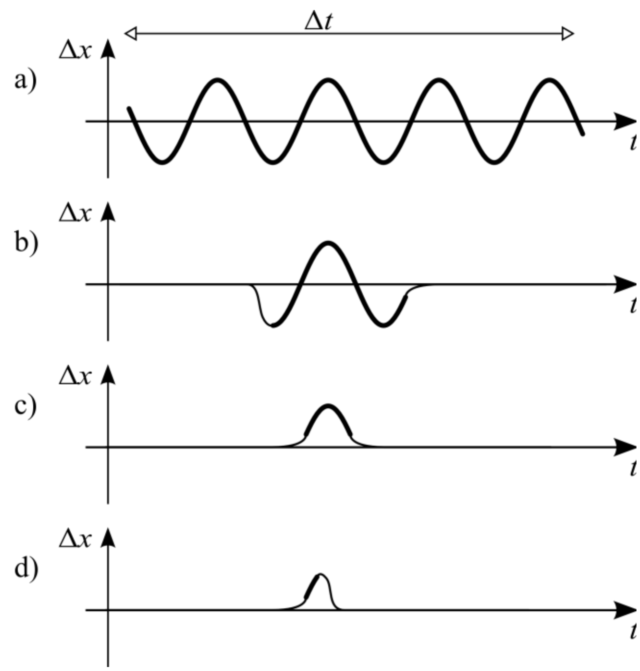


Figure 46.3

Progressively shorter time intervals of the same simple harmonic motion. Bold curves match the sinusoid exactly, thin curves provide continuity to $\Delta x = 0$.

For a sample of sound with a known duration, this means that you can never know any contributing frequencies precisely. The minimum possible uncertainty is called the **frequency resolution** for that sample, given by

$$\Delta f = \frac{1}{\Delta t} \quad . \qquad (46.2)$$

Sadly, the same symbol is typically used for both frequency uncertainty and frequency resolution.

On a spectrogram, Δf and Δt are intervals along the two axes, and their product defines an area in a spectrogram. The Gabor limit means that there is a smallest size "pixel" that can be used to specify what the spectrogram looks like. Figure 46.4 illustrates this. It shows the spectrogram of a tone with four partials, each of which rises in frequency over a duration of about 0.2 s. The frequencies involved are near the midrange for human perception. Part (a) shows an idealized version, which can be described clearly enough, but which would be impossible to actually create. Part (b) shows a spectrogram of a real vibration or sound that approximates the ideal. The rectangular patches have an area

$$\Delta f \, \Delta t = (50 \text{ Hz})(0.02 \text{ s}) = 1 \quad . \qquad (46.3)$$

The same exact sound can be analyzed with a better resolution in the frequency, but only by analyzing with a worse time resolution. Part (c) shows an example, where the patches have area

$$\Delta f \, \Delta t = (25 \text{ Hz})(0.04 \text{ s}) = 1 \quad . \qquad (46.4)$$

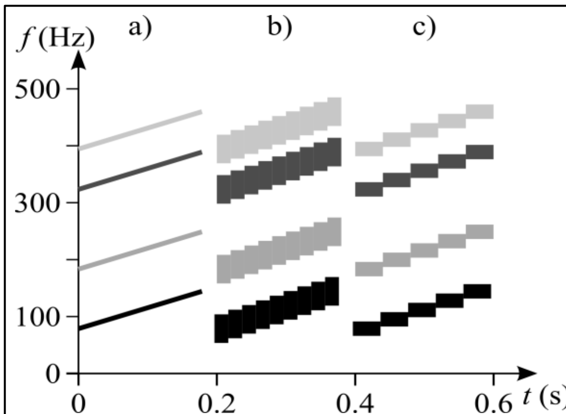


Figure 46.4

Spectrogram of sound with four partials and rising pitch. (a) Impossibly precise ideal. (b–c) Realizations of ideal that obey Gabor limit.

While analyzing a sound, there are various **window functions** that can be used to sample different time intervals of the sound and determine the spectrum within those intervals. They have the effect of smoothing out the spectrogram, compared to the jagged appearance of Figure 46.4. But regardless of the method used, the Gabor limit means that it is impossible to precisely identify the frequency for short time intervals.

Chapter 47. Spectrum Superposition

Chapter 37 shows that when two sounds combine, their vibration graphs combine by superposition, which takes quite a bit of effort to do by hand. How do their spectra combine? Really that question has already been answered; we just need to work out the consequences of previously stated rules.

A complex sound is the superposition of several pure tones. But showing the sound's spectrum does not require adding numbers at all. Instead, the peaks for each of the pure tones are all drawn together on one set of axes. Two complex sound spectra superpose in just the same way, as shown in Figure 47.1. Each SHM/partial in each original sound simply becomes a member of a larger set.

> Spectra superpose by combining, not by mathematical addition.

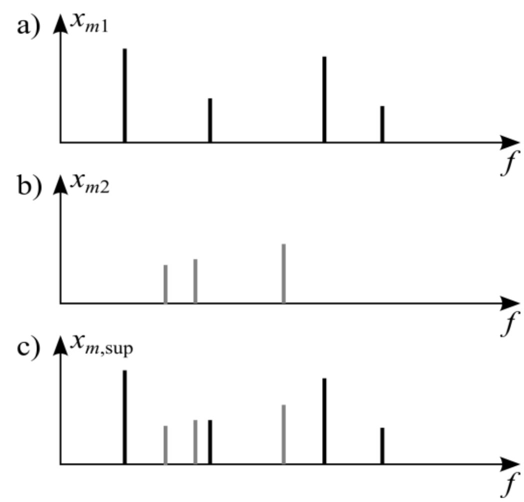


Figure 47.1

Superposition of two spectra. Second spectrum is in gray only to help distinguish them.

But there is one caveat. If two peaks happen to be at exactly the same frequency, then we have the interference situation described in Chapter 39. How those two will combine depends on their phase difference. But unfortunately, all phase information had been removed in order to make the spectra. So the answer cannot be found.

If forced into this situation, the best you can do is to use the average of all the possibilities. Using mathematics beyond the scope of this book, this leads to the following rule:

> When two spectra peaks at the same frequency superpose, on average the resulting peak will be slightly larger than the larger of the two contributing peaks.

(It turns out that even with advanced mathematics, a precise formula for this cannot be obtained, because it involves something called an **elliptic integral.**)

Chapter 48. Chords

Musicians already have a name for making two complex sounds simultaneously: that's a musical **chord.** But Figure 47.1 then reveals a paradox. If a chord is played with three keys on a piano, it can be heard to consist of three parts. And certainly, if two band instruments, say a trumpet and a flute, play simultaneously, then the instruments are distinguishable. But those complex sounds combine in the air, and the partials all enter your ear together. Why would Figure 47.1(c) be perceived as two complex sounds (associated with spectra (a) and (b)), instead of a single combined sound. For that matter, why would we perceive any complex sound as a single entity, instead of as a collection of pure tones?

The answer from psychoacoustics is that it actually can be perceived in any of these ways. When many partials are perceived as a single, complex sound, that's said to be **synthetic listening,** also called **holistic listening.** When a sound is perceived as having separate parts, including hearing each of the partials as an individual sound, that's called **analytic listening.** Different listeners vary in their tendency to one or the other extreme, and one can voluntarily shift perspective, an ability that can be enhanced by musical training. Sounds themselves can shift the listener's perception as well. While a sound is being heard synthetically, if one of the partials changes, especially in frequency, it can often be perceived as splitting out of the complex sound.

There are many subtle factors that influence listeners towards synthetic or analytic perception. One example is linked to duration. In the case of a flute and trumpet, if the flute sound starts a fraction of a second before the trumpet sound, that would help the auditory system (including neural processing) to group the flute partials together, analytically separated from the trumpet partials. If they start at precisely the same time, then the partials are more likely to be synthetically merged. Another major influence is prior experience. Patterns of partials can be recognized as arising from a single recalled source, and are thereby separated out of the full complex sound. One of the most powerful patterns in this respect is a harmonic sequence, presumably because periodic vibrations are so common in our environment. So, if a harmonic set of partials can be identified as part of a complex vibration, they are more likely to be synthetically perceived as a single sound.

Chapter 49. Power

Sound carries energy through the air (or whatever medium it is traveling through). In Chapter 29 energy is described as something possessed by a system. But to understand energy transmission, we need to change focus from the energy *in* a system to the energy exchanged *between* systems — that is, the work.

Figure 49.1 illustrates a speaker producing sound, which travels down an open-ended tube to a listener. The vibrating speaker has energy, and it can affect the air. The result of that action is to transfer energy to the air (that is, to do work on the air), where it becomes sound energy. That sound energy is carried to the other end of the tube, as each parcel of air does work on its neighbor. At the end, the air does work on the listener's ear, and the transferred energy changes from sound back into vibrational energy.

Part of the strength of the energy perspective is that we can analyze this picture without knowing anything additional about what kind of energy is in sound. Be it kinetic or potential, spring-like or oscillating or anything else, all we need to know is that it is conserved.

As long as the speaker continuously emits sound, the total amount of energy transferred keeps increasing continuously. Without knowing when the sound starts and stops, we can't know the total amount transferred. But if the energy transfer is steady, then the energy ΔE transferred in any chosen time interval Δt is proportional to that time interval. Thus, we are led to the ratio called the **power,**

$$W = \frac{\Delta E}{\Delta t} \quad . \qquad (49.1)$$

In physics, the traditional algebraic symbol for power is P, but in this book we will use the usual acoustic choice of W. The SI root unit for power is the $\mathrm{watt} = \mathrm{W} = \mathrm{J/s}$, named after the 18th century Scottish engineer James Watt. Whether it is a good or bad idea to have the algebraic symbol match the unit symbol, I leave to your opinion.

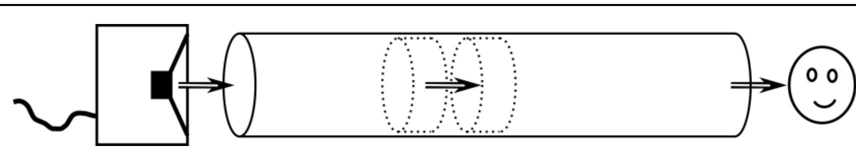


Figure 49.1

Energy transferred from speaker, down open tube of air, and to listener's ear. Double-line arrows represent energy transfers.

Power can describe a few subtly different situations. If the sound source in Figure 49.1 were a tuning fork instead of a speaker, then once some energy transfers from tuning fork to air, the energy in the tuning fork is depleted. The ΔE in Eq. 49.1 would equal the reduction in the tuning fork's energy, giving the Δ the usual meaning of "change in …" From this perspective, the power W would be the rate of energy loss for the tuning fork.

On the other hand, the energy contained by a speaker driven by an amplifier does not change, because it is continuously re-supplied through the wires. But for both the tuning fork and the speaker, ΔE can be considered the energy transferred from the source into the air, and W is the rate of energy transfer.

As the sound energy travels down the tube, it is transferred from one parcel of air to the next, as represented by the middle double arrow in Figure 49.1. Here there are not separate objects that the energy is being transferred from or to, but the energy does move. A third interpretation of power is thus as a rate of energy transmission through space, rather than a transfer between objects.

Chapter 50. Intensity

Figure 50.1(a) modifies the situation in Figure 49.1 by adding another, smaller, open-ended tube that the sound must pass through in order to get to the listener. As the sound transitions from the larger to the smaller tube, much of it heads elsewhere, along with its power; the power captured depends on the area of the small tube opening. Conversely, Figure 50.1(b) shows a listener using an ear trumpet, a hearing aid used in the days before electronics. The large circular area of the opening of the ear trumpet collects more power than the ear alone would, and that power is then funneled into the ear.

These examples illustrate that the power of a sound is spread over an area. Indeed, if the power is spread uniformly, then in each of these examples the power W collected is proportional to the area A used. For example, if the end of the smaller tube in Figure 50.1(a) has half the area of the end of the larger tube, then we would expect it to "catch" half of the sound power coming down the larger tube. We are thus led to the ratio called the **intensity,**

$$I = \frac{W}{A} \quad . \tag{50.1}$$

The SI unit for intensity, directly determined from this ratio, is $\mathrm{W/m^2}$, which does not have a special name.

It is important here that the area used is a **cross-sectional area,** that is, an area that is perpendicular to the direction in which the sound is moving. If the ear trumpet were not pointed directly at the sound source, then less power would enter it. The calculation of Eq. 50.1 would then yield a smaller value, even though the sound itself is unchanged.

It's not necessary for a physical object to define the area. An imaginary area will serve just as well, as long as it is cross-sectional. If the gentleman with the ear trumpet leaves, while the speaker continues to sound, the sound intensity at that place is the same as it ever was. And because of the proportion, the same intensity will be calculated for any cross-sectional area that might be chosen, as long as we use the corresponding power.

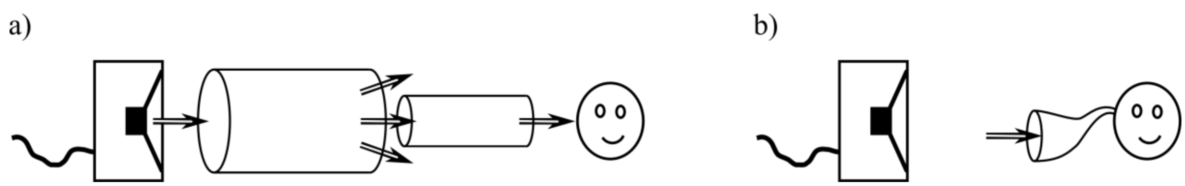


Figure 50.1
Examples where amount of energy transmitted or transferred depends on a cross-sectional area.

As a sound with total power W travels, it may spread over different areas. For example, after entering the ear trumpet in Figure 50.1(b) the sound energy is funneled from a larger to a smaller area. Energy conservation assures us that as this happens, energy is not lost. Since energy enters the ear trumpet at the same rate that it leaves, this means that the sound power entering through the larger area is the same as the sound power leaving through the smaller area. This leads to the relationships

$$I_1 A_1 = W_1 = W_2 = I_2 A_2 \quad , \tag{50.2}$$

$$\frac{I_1}{I_2} = \frac{A_2}{A_1} = \frac{1/A_1}{1/A_2} \quad , \tag{50.3}$$

$$I \propto \frac{1}{A} \quad . \tag{50.4}$$

In summary, for constant power (such as when the same sound propagates so that its cross-sectional area changes), the intensities are **inversely proportional** to the areas that the sound is passing through or impacting. Focusing on the first two members of Eq. 50.3, notice that for an inverse proportion the subscripts are twisted out of their usual arrangement. For example, compare to the subscripts in Eq. 4.23, an example of a direct proportion.

Chapter 51. Range of Intensity Perception

In Chapter 36 it is mentioned that vibrations with larger amplitude create louder sounds. But source amplitude is not a good way to measure sound loudness, because other factors are also important. For instance, loudness of a sound can depend on the distance from source to listener. Perhaps a better measure of sound loudness would be the amplitude of some responding vibration, for instance in the listener's ear. But that has the problem that it would be different for different sound detectors; your ear and a microphone might respond to the same sound with different vibration amplitudes. For sounds, intensity is a better measurement of loudness.

The smallest intensity sound that humans can hear is called the **threshold of audibility** or the **threshold of hearing.** A standard value for this is $I_0 = 10^{-12}\,\mathrm{W/m^2}$, although there is variation among individuals and for different types of sounds, not to mention a natural degradation of hearing with age. The symbol I_0 is used for this specific value, not the variable threshold itself. (See Chapter 64 for how the audibility threshold depends on frequency.)

This is an amazingly small intensity. If this were in light energy instead of sound energy, this would be the equivalent of being able to see a 4 W night-light from a distance of about 350 miles away! To hear such quiet sounds of course requires optimal conditions. Not only must the person have hearing that is not degraded, but they must spend time in a quiet environment to become sensitized. During normal living, background noises are rarely less than $10^{-9}\,\mathrm{W/m^2}$.

On the other end of the scale, very loud sounds can cause discomfort, louder sounds can cause a tickling feeling, and louder still can cause physical pain. The intensities which are just large enough to cause each of these sensations are again called thresholds. Compared to the threshold of audibility, they all vary less with frequency (see Chapter 64), but they vary more depending on the individual involved and on environmental conditions. For instance, the **threshold of discomfort** changes the longer a person is exposed to loud sounds. Despite this variability, a standard value for the **threshold of discomfort** is $1\,\mathrm{W/m^2}$. This is again a relatively small intensity, the equivalent in light energy being a 4 W night light held at an arm's length.

The **threshold for damage** is lower than the **threshold of pain.** Sounds that are not physically painful can still do permanent damage to the hearing apparatus. When considering damage, duration of exposure is a factor as well as intensity, which is reflected in sound safety standards.[10]

Human hearing is thus quite sensitive compared to vision. But even more impressive is the range between the audibility and pain thresholds. Of the human senses, only vision encompasses a larger range of stimulus strength. Example sounds for different intensities are readily available on the Internet. A good example is at *Noise Help*.[11]

Chapter 52. Limitations of the Intensity Model

Intensity, described in Chapter 50, works very well to describe the transport of energy by sound when you know the paths along which the sound is traveling. But it is not a perfect model for describing the loudness of sounds as they are heard or detected. One way to describe the difficulty is that when sound meets an obstacle, it may not travel in a simple path.

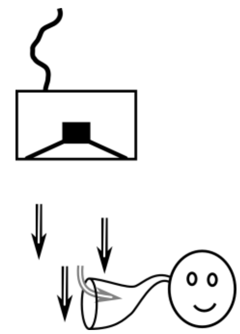


Figure 52.1
Sound traveling around corners. Intensity alone doesn't work well to describe the loudness perceived.

Based on the descriptions of Figure 50.1(b), you might think that in Figure 52.1 the listener wouldn't hear anything, because the ear trumpet is oriented so that it doesn't "catch" any of the sound. However, as mentioned in Chapter 8, sound can travel around corners. Sound energy would enter the ear trumpet in this second situation. Indeed, you know from personal experience that your ear does not need to be pointing directly at a sound source in order to hear it.

To make improvements to the model would require getting into the details of how sound moves through the medium, instead of relying on the more general properties of energy conservation. We will not enter into those details at this point. In any case, the intensity model remains very useful, whenever one wants to describe how loud a sound is without worrying about exactly how the listener is tilting their head.

Chapter 53. Intensity and Distance

53a. Area and Length

You are probably familiar with formulae to calculate the areas of various particular shapes. Table 53.1 lists some such formulae, for four flat shapes and the outer surface areas of two 3D shapes. But there is also relationship that is valid for **any** shape, a relationship between different areas of an arbitrary shape and the different lengths you might measure on it. This relationship is a property of geometry, the subfield of mathematics, but it has many applications in physics. We will apply it to sound intensity in Section 53b.

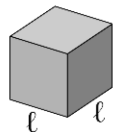
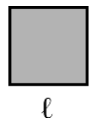
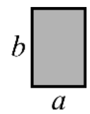
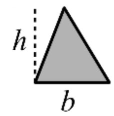
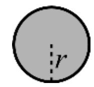
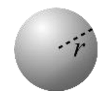
We start with the very definition of measurement. At the most basic level, to measure something you first choose a unit quantity, and then count how many of the unit quantities are required to equal the thing being measured.

For example, to measure the area of the blob in Figure 53.1(a), we might choose the light gray square as a unit area. It doesn't have to be a square, but squares are nice to work with. It might also be nice to know the linear dimensions of the unit area; if the square is 1 cm on each side, then we could call the unit area 1 cm^2. But that is not required. The fundamental thing is that we have a gray patch that is our unit area.

[10] "Appendix II:A. General Industry Standard," *Noise and Hearing Conservation,* ed. Brian Liddell, in *OSHA Technical Manual,* n.d., https://www.osha.gov/dts/osta/otm/noise/standards_more.html (July 2013)
[11] Sarinne Fox, "Noise Level Chart," *Noise Help,* n.d., http://www.noisehelp.com/noise-level-chart.html (July 2013)

Table 53.1

Area formulas for specific shapes.

Cube	Square	Rectangle	Triangle	Circle	Sphere
					
$A = 6\,\ell^2$	$A = \ell^2$	$A = ab$	$A = \frac{1}{2}\,hb$	$A = \pi\, r^2$	$A = 4\pi\, r^2$

Replicating that unit square creates a grid, from which it appears that about 24.5 of those unit squares are required to cover the blob. Arriving at that number requires quite a bit of estimating at the blob edges, where only fractions of the unit square are inside the border. To get a more accurate measurement, the simplest option would be to start with a smaller unit area. There are other, more sophisticated methods of getting more accurate results, but they are all refinements on the fundamental idea that (roughly) 24.5 unit squares are required to cover the blob.

Figure 53.1(b) shows exactly the same shape, except that it has been scaled up, or magnified, by a factor of two. The technical term in geometry is that the two blobs are **similar**. The "factor of two," called the **scale factor,** means that the width in (b) is twice the width in (a), and the height of (b) is also twice the height of (a). In fact, every linear measure of (b) is twice that of (a): the length of the boundary, the distance between the two bumps on the left, the width of the square-ish bit on top, … everything. This may not be obvious, but it is true.

> When a shape is scaled uniformly, all linear measures (lengths and distances) on that shape are proportional to all other linear measures. The ratio of a length on the new shape to the corresponding length on the original shape is called the **scale factor.**

Figure 53.1(b) also shows the measurement grid similarly magnified. However, those larger squares are not unit squares; the unit of measurement remains the gray square in Figure 53.1(a). In fact, the larger grid squares have an area equal to $4 = 2 \times 2 = 2^2$ unit squares. Since the blob and grid were both magnified

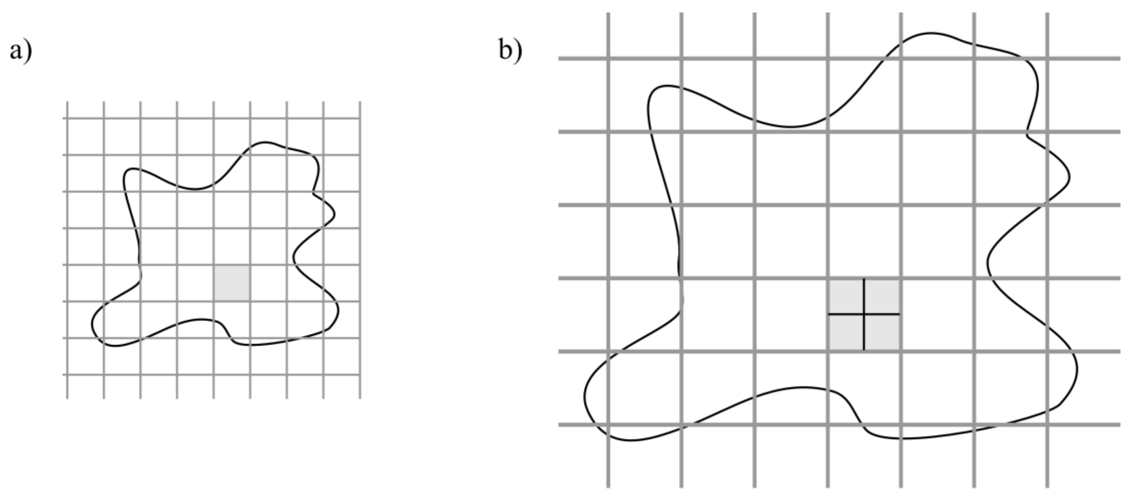


Figure 53.1

Area measurement of two arbitrary, similar (in the technical sense) shapes. Shape (b) has double the scale of shape (a).

in the same way, the large blob is covered by the same number of larger squares. Therefore, the area of the large blob must be 2^2 times the area of the small blob, or roughly $24.5 \times 4 = 98$ unit squares.

This relationship holds for any scale factor. If the scale factor were 3, then each scaled grid square would have an area of $3^2 = 9$ unit squares.

> When a shape is scaled by a factor S, the any area associated with that shape changes by a factor S^2.

Notice that the argument above would work for any shape. In fact, it would work for any area associated with that shape: the total area, the area of the lower left bump, etc. The rule even works for three-dimensional shapes like balls and boxes. When the scale factor is not a whole number, it is less obvious that the same rules apply, but they do.

Combining the two above principles, the way that areas scale can be related directly to lengths.

> If two shapes are geometrically similar, then any area associated with the shape is proportional to the square of any length on the shape.

$$A \propto l^2 \quad , \tag{53.1}$$

Keep in mind that this rule only works for two shapes that are similar in the technical, geometric sense. For instance, a tuna can and a soup can are not similar in this sense. Even though both are cylinders, their ratios of height to diameter are not the same.

53b. Inverse Square Law

Consider a very brief sound, like a clap, with a duration Δt. Energy E is transferred from the source into the air, and then moves away from the source in a shell. Figure 53.2 shows this in a situation where the clap was done very close to the ground, so that the sound cannot travel downwards. Modelling the source as a point source, the shell therefore has the shape of a hemisphere. In part (a) the end of the sound (which is the **T**railing edge) has traveled r_{Ta} from the point source, while the beginning of the sound (which is the **L**eading edge) has traveled a little further, $r_{La} = r_{Ta} + s\,\Delta t$, because it had a head start. In part (b) the sound has traveled further from the point source, but the thickness of the shell $r_{Lb} - r_{Tb}$ is the same, since both the leading and trailing edge are moving at the same speed.

As the sound travels very little of the sound energy is converted to other forms, such as turbulence in the air. For sounds traveling less than a few miles, it is a good model to ignore such conversion, so that energy conservation implies that the same energy E is contained in the shells of both Figure 53.2(a) and (b).

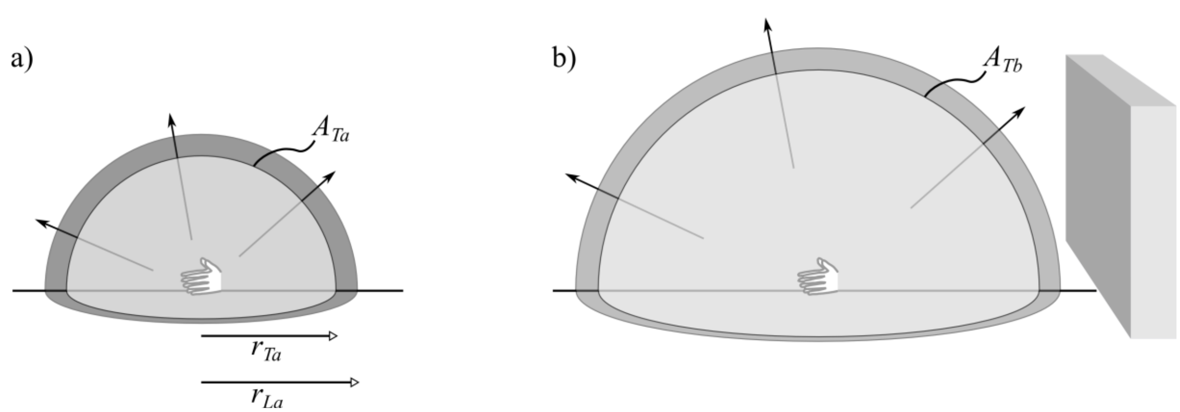


Figure 53.2

As sound travels away from a point source, sound power gets spread over larger area, leading to lower intensity.

However, in part (b) that energy has been spread over a larger volume, which is why the sound is softer at that larger distance.

During the time Δt immediately before Figure 53.2(a) all of the energy passed through the hemisphere with radius r_{Ta}. So, the power passing through that complete hemisphere is $W = E/\Delta t$, the same rate as the power at which the sound was created. That power is spread over the hemispherical area A_{Ta}, so that the intensity there is $I = W/A_{Ta}$. By the later time in part (b), the power is spread over the larger area A_{Tb}. Those shapes (hemispheres) are geometrically similar, so that from Eq. 53.1 comes the proportion

$$\frac{A_{Tb}}{A_{Ta}} = \left(\frac{r_{Tb}}{r_{Ta}}\right)^2 \quad . \qquad (53.2)$$

Since the intensity is inversely proportional to area, as in Eq. 50.4, the intensities are related by

$$\frac{I_{Tb}}{I_{Ta}} = \frac{A_{Ta}}{A_{Tb}} = \left(\frac{r_{Ta}}{r_{Tb}}\right)^2 \quad , \qquad (53.3)$$

$$I \propto \frac{1}{r^2} \qquad (53.4)$$

This **inverse-square relationship** shows up quite often in physics, whenever there is a point source of things that travel at a known speed; it could be light, or radioactive particles, or other things besides sound.

If you happen to know the formula for the area over which the power is spread, then the intensity can be found as a function of distance from the source. For Figure 53.2, the case of a hemisphere, we get

$$I = \frac{W}{2\pi r^2} \quad , \qquad (53.5)$$

where W is both the power passing through any hemisphere and the power leaving the source. Another commonly encountered situation is when the sound is leaving the source equally in all directions. The sound then forms a sphere around the source, so that

$$I = \frac{W}{4\pi r^2} \quad . \qquad (53.6)$$

When a sound departs its source in a set of directions that do not produce a simple shape, you may not know how to calculate the shape's area. But as long as the shape stays geometrically similar, proportion 53.4 still works as a starting point. Recall that the way to use a proportion is to create an equation from it. To relate the intensities at two different distances, we can write

$$\frac{I_1}{I_2} = \frac{1/r_1^2}{1/r_2^2} = \frac{r_2^2}{r_1^2} = \left(\frac{r_2}{r_1}\right)^2 \quad . \qquad (53.7)$$

Even if the shape of the sound surface does not stay geometrically similar, the inverse square relationship *may* be applicable. In Figure 53.2(b) the sound is about to meet a wall. If the sound energy that hits the wall is absorbed, then that will have no effect on the sound that is traveling in other directions. So Eq. 53.4 and 53.5 will still be accurate, for instance, on the side away from the wall. On the other hand, if some of the sound reflects off the wall, then intensities in other locations may be affected and those equations will no longer apply, at least not so simply.

Perhaps the most extreme example of this is inside a room. Unless the walls have been specially made to absorb sound, forming an **anechoic chamber,** then at least a portion of the sounds hitting the walls will reflect back into the room. For this reason, the inverse-square relationship is of limited use in considering the acoustics of a room.

53c. Extra: Not the Inverse Square Law

If a sound source cannot be well modeled as a point source, this will also change the intensity-distance relationship. For example, Figure 53.3 shows a river creating the rushing sound of the water. As that sound moves away from the river, it keeps the shape of a half cylinder. However, the shape does not stay similar in the technical geometric sense, because the length stays the same while the radius increases.

In these situations, useful relationships may still arise from the relationship between area and distance. But they will not be the inverse square relationship. Cases like this are not detailed in this book.

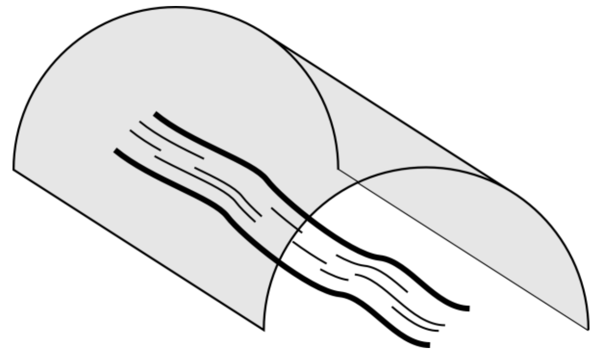


Figure 53.3
Sound from a river.

Chapter 54. Fechner's Law

In the study of perception, there is a rule of thumb called Fechner's Law:[12]

> Human sensations are logarithmically related to the physical quantities that they sense.

The word "logarithmically" means that equal increments in perception correspond to equal *multiples* of the physical quantity, instead of equal increments.

For example, on a music player volume knob that has evenly spaced marks or numbers, incrementing the knob by one mark usually corresponds to *multiplying* the output sound power of the stereo by some factor. (In case you haven't read Chapter 49, power is a physical measure of "how much" sound the music player is producing.) Suppose that increasing the knob from 1 to 2 results in doubling the sound power. Then increasing from 2 to 3 will double it again, meaning that 3 on the knob produces four times as much sound power as 1 on the knob. Increasing from 3 to 4 will double the power yet again, to eight times the power of 1 on the knob.

Fechner's Law applies, with varying degrees of accuracy, to all of the senses, in some cases including multiple aspects of one sense.

Chapter 55. Logarithms

55a. Defining the Logarithm

Fechner's Law is not a proportional relationship. Algebraically, the relationship is expressed with the logarithm function, log for short. The logarithm is the inverse operation of raising a number to a power. In fact, one description is that a logarithm *is* an exponent, defined by the equations

$$x = 10^X \quad \Leftrightarrow \quad \log x = X \quad , \qquad (55.1)$$

where capitalized variables are being used for exponents. This relationship is graphed in Figure 55.1. Measurements of perception (like marks on a volume knob) show up in equations placed the way X is in Eqs. 55.1, and measurements of the physical stimulus (like the power from a stereo) show up in equations placed the way x is. The number 10 in Eq. 55.1 is called the **base** of the logarithm. In the volume knob example in Chapter 54, the doubling of the sound power suggests using a base of 2. But it turns out that

[12] Arthur S. Reber, *The Penguin Dictionary of Psychology*, 2nd ed. (London: Penguin Books, 1995), 283.

the choice of base is only a matter of preference or convenience. Throughout this book base 10 will be used for all purposes.

In some uses of exponents, such as scientific notation, we choose to only use integers. The meaning of integer exponents can be described as repeated multiplication. But the variable X need not be an integer. Fractional exponents cannot be described as repeated multiplication, but they do work mathematically; for instance, an exponent of 0.5 means the same thing as a square root. Other fractional numbers in the exponent may be harder to explain, but your calculator can calculate them.

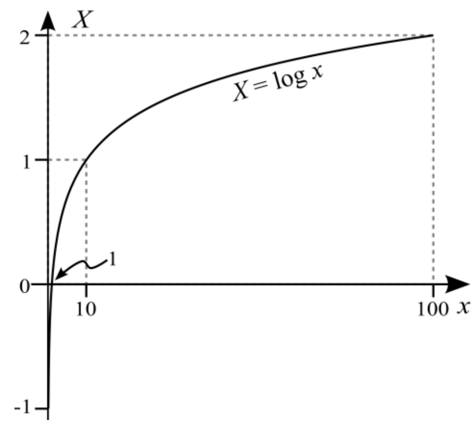


Figure 55.1
Graph of logarithm function.

The logarithm of a negative number is not defined. In the first element of Eq. 55.1, there is no X that would make $x < 0$. That is why in Figure 55.1 the function curve stays to the right of the origin. The logarithm X can have either sign, however. For any logarithm base, $\log(1) = 0$, so that a positive logarithm X means that the original number x was greater than one, and a negative logarithm X means that the original number x was between 0 and 1.

Because logarithms are exponents, they obey several special rules.

$$10^X 10^Y = 10^{X+Y} \quad \Leftrightarrow \quad \log(xy) = \log x + \log y \tag{55.2}$$

$$\frac{10^X}{10^Y} = 10^{X-Y} \quad \Leftrightarrow \quad \log\left(\frac{x}{y}\right) = \log x - \log y \tag{55.3}$$

$$(10^X)^a = 10^{a\cdot X} \quad \Leftrightarrow \quad \log(x^a) = a \log x \tag{55.4}$$

The righthand equalities above can be handy for manipulating equations containing logarithms, without necessarily going back to how they derive from the lefthand equalities.

55b. Extra: Interpreting Logarithms

Base-10 logarithms can be interpreted by looking at the integer and fractional parts separately. If $X = \log x$ is positive, which implies that $x > 1$, then we can break up the logarithm as

$$\log x = X = A.B = A + 0.B \quad , \tag{55.5}$$

where A and B represent the digits in X on the two sides of the decimal point.

- The fractional part B determines what x has for significant digits, regardless of where the decimal point is in x. Figure 55.2 is a scale representing all possible fractional parts B on the right side. The left side represents all possible sequences of significant digits. Since a leading zero can't be significant, the smallest number on the left side is 1.

- The whole number part A indicates the position of the decimal point in x.
  - There are $A + 1$ significant digits to the left of the decimal point in x.
  - In other words, starting with the x value from the left of Figure 55.2, A is the number of places to move the decimal point towards the right.
- Examples:
  - $X = \log x = 4.5 \rightarrow B = 0.5$ means that the digits in x are 3.162. $A = 4$ moves the decimal to $x = 31620$.
  - $X = \log x = 2.845 \rightarrow B = 0.845$ gives digits 7.0 $A = 2$ moves the decimal to give $x = 700.0$.

In this way, multiplying (or dividing) the original number x by 10 (that is, shifting the decimal point by one digit) corresponds to increasing (or decreasing) A by one in the logarithm.

If the logarithm is negative ($X < 0$), which implies that $0 < x < 1$, then a similar situation applies. Dividing the x values on the left of Figure 55.2 by a power of 10 still corresponds to subtracting a whole number from the logarithm. The difference is that $0.B - 1$, which is negative, has different fractional digits than $0.B$. For example, if $0.B = 0.63$, then $0.B - 1 = 0.37$. Calling those new digits D, we have the relationships

$$\log x = X = -C.D = 0.B - (C + 1) \quad , \qquad (55.6)$$

$$0.B = 1 - 0.D \quad . \qquad (55.7)$$

- The fractional part D of the negative logarithm still determines what x has for significant digits. Find them by first applying Eq. 55.7 and then finding the resulting B in Figure 55.2.
- C is the number of zeros in x between the decimal point and the first significant digit.
- Examples:
  - $X = log\, x = -1.0969 \rightarrow D = 0.0969$, which implies $B = 1 - 0.0969 = 0.9031$. Finding this on the scale shows that the digits in x are 8.0. The integer part of the logarithm has $C = 1$, which positions the decimal place so that $x = 0.080$.
  - $X = log\, x = -3.3 \rightarrow D = 0.3$, which implies $B = 1 - 0.3 = 0.7$. The scale shows the digits in x to be 5.011. The integer part of the logarithm has $C = 3$, which positions the decimal place so that $x = 0.0005011$.

Changing the whole number part C by one still corresponds to multiplying (or dividing) x by 10, just as with the whole number part A in the positive logarithm case.

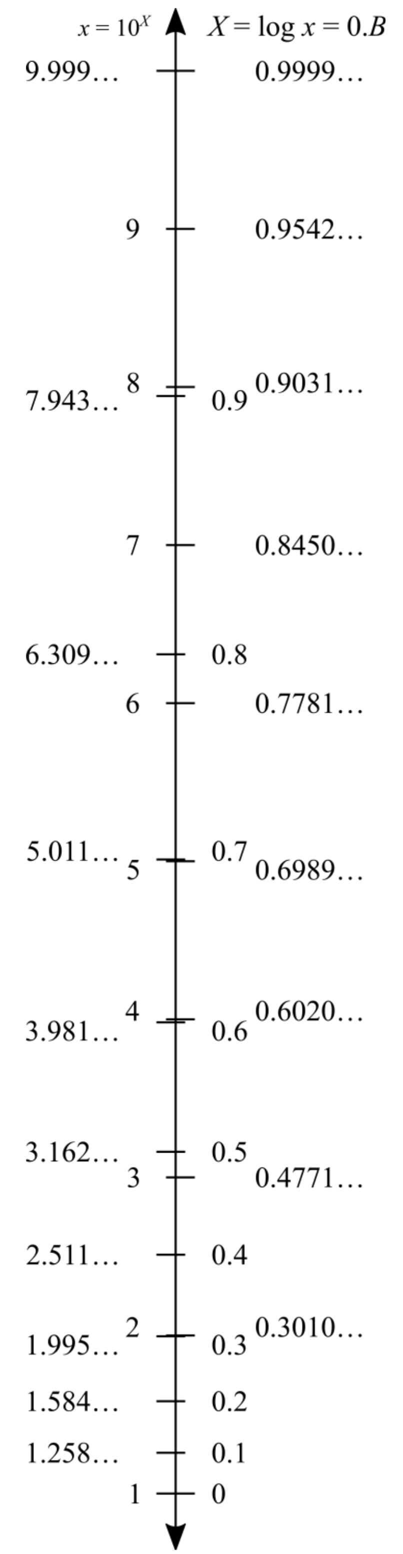


Figure 55.2

Scale with numbers (left) and their logarithms (right). Numbers with one significant digit are exact.

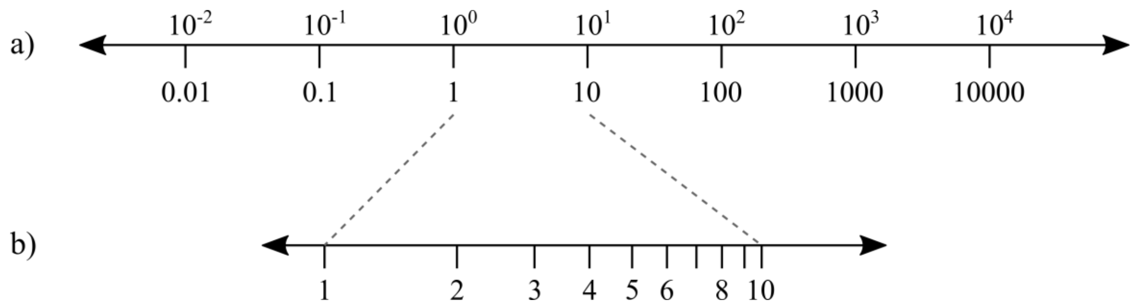


Figure 56.1
Logarithmic axes.

Chapter 56. Logarithmic Axes

When visually comparing numbers that cover a very large range of sizes, it can be helpful to use a **logarithmic axis.** This is an axis where equal steps along the line correspond to equal increments in the logarithm of the number represented; position on the line is proportional to *the logarithm of* the number represented.

Even though the logarithm determines how to arrange the numbers on a logarithmic axis, such an axis can be used without ever considering the logarithm function. An example is given in Figure 56.1. Just think of it as an axis where equal sized steps along the line correspond to equal *multiples* of the represented quantity. In Figure 56.1(a), it is evident that moving from one label to the next corresponds to multiplying by ten. What is not so obvious is that this is true no matter where you start. If you find 23 on the axis, and move that same distance to the right, you arrive at 230. Starting at 0.5 and moving that distance to the left takes you to 0.05. Even more amazing is that the same sort of thing works, on this same axis, for any multiplier. Even though Figure 56.1(a) has been labeled in a way suggestive of base-10 logs, there exists on that axis a step size that corresponds to multiplying by 2, or 14, or any other factor.

If the marks are labeled using exponents, as on the top side of Figure 56.1(a), then a resemblance to a linear axis becomes apparent: the exponents just count up and down, as you often see on a linear axis. But it is critical to keep in mind that this similarity is something of an illusion. A step along the axis still corresponds to multiplication, and the exponents are just making that look like addition.

When estimating values for positions between marks, logarithmic axes are a particular challenge. As illustrated in Figure 56.1(b), the regular counting steps are not equally spaced. As the numbers increase, the intervals get smaller. Nothing in this book will require you to make such estimates with precision, but a handy estimate is that the step for a factor of 2 is very close to three-tenths (that is, a bit less than one third) of the step for a factor of 10. In Figure 55.2, notice how close $x = 2$ is to $X = 0.3$.

As shown in Figure 56.1(b), is it common to skip labels as the tick marks get closer together. In fact, occasionally axes are drawn with some of the tick marks themselves left out. In those cases, the reader is expected to fill in the missing information.

Notice that a logarithmic axis can never show negative numbers. Moving to the left, dividing with each step, leads to smaller and smaller numbers, but can never even reach zero.

Chapter 57. Sound Intensity Level

If you have read Chapter 91, you will notice that this subject very much parallels that chapter.

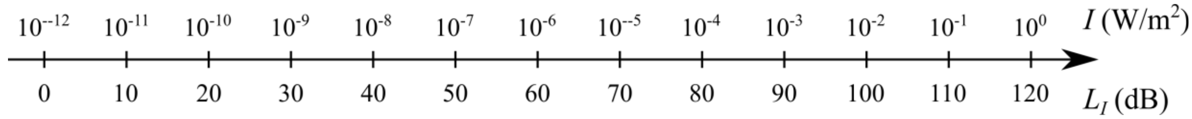


Figure 57.1

Axis showing relationship between Sound Intensity Level and intensity over the human range of perception.

Moderately loud sounds, such as conversation, have intensities near $10^{-6}\,\mathrm{W/m^2}$. Notice that this is nowhere near the arithmetic mean of the thresholds of discomfort and audibilty,

$$(1\,\mathrm{W/m^2} + 10^{-12}\,\mathrm{W/m^2})/2 = 0.5000000000005\,\mathrm{W/m^2} \approx 0.5\,\mathrm{W/m^2} \quad .$$

As predicted by Fechner's Law, it is instead the exponents (that is, the logarithms of the intensities), that should be averaged to get to moderate: $\big(0 + (-12)\big)/2 = -6$. Because of this relationship to perception, graphs of sound intensities nearly always use logarithmic axes, in order to conveniently display a reasonable array of different loudness perceptions.

This is one of the main motivations for defining a different way to specify intensity called **sound intensity level**, or **SIL**. The usual algebraic symbol for SIL is L_I, although β is also commonly used. It is measured in units of decibels (abbreviation dB). The SI root unit here is the **bel**, and **deci-** is a metric prefix that is too unimportant to make it into Table 4.1. But it is extremely uncommon to use anything but the combination dB.

Figure 57.1 illustrates how sound intensity levels are related to intensities over the range of human hearing. As one moves along this axis, each tick mark corresponds to the intensity increasing by a *factor* of ten (a logarithmic scale), while sound intensity level increases by an *increment* of 10 dB (a linear scale). Keep in mind that SIL and intensity are two ways of measuring the same thing, somewhat the way centimeters and meters are two ways to measure length. The conversion between them is more complex, though, using the equations[13]

$$L_I = (10\text{ dB})\log\left(\frac{I}{I_0}\right) \quad , \tag{57.1}$$

$$I = I_0 10^{(L_I/10\mathrm{dB})} \tag{57.2}$$

$$I_0 = 10^{-12}\,\frac{\mathrm{W}}{\mathrm{m^2}} \tag{57.3}$$

These equations can be thought of as the prescriptions for moving from one side to the other of the scale in Figure 57.1.

Each of the words in "SIL" means something specific. "Level" denotes something that has units of decibels and that is defined by a logarithm equation like Eq. 57.1. There are other "level" quantities, which differ from SIL only by what is inside the logarithm function; the word "Intensity" specifies that. The word "Sound" indicates not only the type of energy under consideration, but also what value to use for I_0. The value in Eq. 57.3 is chosen by convention, and it does not have any physical significance. It does have significance for humans though, as it is the same standard **threshold of hearing** defined in Chapter 51.

Notice that zero plays a different role on the two scales. An intensity of zero would indicate no sound whatsoever. But because intensity is on a logarithmic scale in Figure 57.1, zero intensity doesn't even appear. If you try to calculate the SIL for $I = 0\,\mathrm{W/m^2}$, your calculator will probably give an error; based

[13] For the purist: The dB unit is not properly part of the (10dB) appearing in these equations. In this context, the result of the logarithm function already has the unit bel, and the 10 arises from the deci- prefix alone. But writing the equations using (10dB) provides an explicit reminder for how the units come out.

on the scale, we might call that $-\infty$ dB. On the other hand, an SIL of $L_I = 0$ dB doesn't indicate no sound; instead, negative SIL indicates a sound level that is too quiet to hear (using the rough model that I_0 is the threshold of hearing).

Another "level" quantity is sound pressure level, or SPL. Pressure is a physical quantity that varies when sound passes through the medium, which is described in Chapter 125. The advantage of considering pressure instead of intensity is that it provides a way to deal with the limitations discussed in Chapter 52. However, those details are beyond the scope of this book. In simple situations, such as sound traveling in straight lines, SPL is defined so that it's the same as SIL. SPL is mentioned here only so that you can recognize it as a synonym for SIL in such situations.

Chapter 58. SIL Comparison

If you have read Chapter 92, you will notice that this subject very much parallels that chapter.

Numerical comparisons can be made in two basic ways: differences (e.g., "This Halloween I got 20 more pieces of candy than my little brother.") and ratios (e.g., "This Halloween I got twice as much candy as my little brother.") Notice that difference comparisons require the choice of a unit (*pieces* of candy in the example), while ratio comparisons do not. One of the benefits offered by level measurements such as SIL is that they turn ratio comparisons into differences.

Suppose we are comparing the sound intensity reaching us from two engines, in a car and in a truck. Comparing their SILs as a difference, we find that

$$\begin{aligned} \Delta L_I = L_{It} - L_{Ic} &= (10\text{ dB})\log\left(\frac{I_t}{I_0}\right) - (10\text{ dB})\log\left(\frac{I_c}{I_0}\right) \\ &= (10\text{ dB})\left[\log\left(\frac{I_t}{I_0}\right) - \log\left(\frac{I_c}{I_0}\right)\right] \\ &= (10\text{ dB})\log\left(\frac{I_t/I_0}{I_c/I_0}\right) \quad , \end{aligned} \tag{58.1}$$

$$\Delta L_I = (10\text{ dB})\log\left(\frac{I_t}{I_c}\right) \quad , \tag{58.2}$$

where the special property of logarithms Eq. 55.3 has been used. The ratio of the intensities translates to a difference in the SILs. Notice how similar Eq. 58.2 is to Eq. 57.1. In fact, one might say that *all* SILs are comparisons, with Eq. 57.1 really being a comparison to the standard I_0.

Level quantities are so directly tied to difference comparisons, that it is *never* appropriate to multiply or divide any quantity with decibel units. (The only exception is the operation with 10 dB when converting between decibels and their underlying quantity, as in Eqs. 57.1 or 57.2.)

When it is appropriate to compare or combine intensities by addition or subtraction, however, SILs become decidedly non-intuitive. Suppose that the intensities in our example are $I_c = 2.0 \times 10^{-6}\text{ W/m}^2$ and $I_t = 8.0 \times 10^{-6}\text{ W/m}^2$, so that $L_{Ic} = 63$ dB and $L_{It} = 69$ dB. If we are exposed to both engines together, the ideas from Chapter 53 tell us that the total intensity will be

$$\begin{aligned} I_T &= 2.0 \times 10^{-6}\text{ W/m}^2 + 8.0 \times 10^{-6}\text{ W/m}^2 \\ &= 1.00 \times 10^{-5}\text{ W/m}^2 \quad . \end{aligned} \tag{58.3}$$

The resulting total SIL is $L_{IT} = 70$ dB, a result which has no intuitive connection to the inputs of 63 dB and 69 dB. In such situations, it is safer to do calculations on the "intensity side" of the scale in Figure 57.1, rather than trying to work with decibels.

Chapter 59. SIL Rules of Thumb

A few easy-to-remember relationships can save having to do calculations with Eqs. 57.1 or 58.2. Whenever intensity is doubled or halved, for instance due to doubling or halving the number of sources, the effect is to change the SIL by ± 3 dB.

Suppose that a sound is expanding away from a point source, so that proportion 53.4 applies. If one starts a distance r_1 from the source and then moves twice as far away to $r_2 = 2r_1$, then the change in the SIL experienced is

$$\begin{aligned} \Delta L_I = L_{I2} - L_{I1} &= (10\ \text{dB}) \log\left(\frac{I_2}{I_1}\right) \\ &= (10\ \text{dB}) \log\left(\left(\frac{r_1}{r_2}\right)^2\right) = (10\ \text{dB}) \log\left(\frac{1}{4}\right) \\ &= -6\ \text{dB} \quad , \end{aligned} \tag{59.1}$$

where Eq. 53.3 has been used to relate the intensity ratio to the distance ratio.

One benefit of working in decibels: In both these cases, the change in SIL is a simple numerical quantity, without needing to specify either the number of sources, or the value of the distances.

Chapter 60. Intensity from Multiple Sources

60a. From One Direction

If a trumpeter and a flutist next to each other play simultaneously, then the combination is louder than when either plays alone. The easiest model for this situation that satisfies conservation of energy is that the energy-related physical concepts obey superposition. The two players are both creating power (i.e., energy per unit time) with their instruments, and all types of energy are comparable. So, the output power of the combination is simply the sum of the two individual contributions. At each place nearby, the sound intensity resulting from the combination is the sum of the intensities that each source would cause individually, expressed in an equation as

$$I_{\text{total}} = I_1 + I_2 + I_3 + \cdots \quad . \tag{60.1}$$

In this model, one sound source does not in any way cancel the sound coming from the other source.

As a special case of this model, if a number N of *equivalent* sources combine, then the resulting power is simply proportional to the number of sources. Similarly, if N sources are situated so that they each cause an intensity of I_1 at a specific location, then the combined intensity is again proportional to the number of sources,

$$I_N = N\, I_1 \quad . \tag{60.2}$$

Note that the location where we are measuring the combination is significant here. It doesn't make sense to talk about "the intensity of a source" without including how the sound reached the place where the intensity is measured.

60b. Extra: From Different Directions

If two simultaneous sound sources are in different locations, then we run into a problem. Chapter 50 emphasizes that it is important to use a cross-sectional area when finding intensity. But an area that is cross-sectional for one of the sounds is not cross-sectional for the other, as illustrated in Figure 60.1. Therefore, it is not quite proper to add together the intensities from the two sources.

Nevertheless, those two sounds do enter the listener's ear, and once they do so they are travelling in the same direction (up the ear canal) and *do* have a common cross-sectional area. So, when considering what the listener hears, the combined intensity *is* the sum of the intensities he would hear from the individual sources. But if the listener turns his head, then how the intensity changes is more complicated. As in Chapter 52, in order to develop a model with fewer caveats we would need to learn how sound moves through the medium.

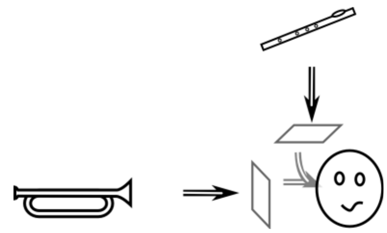


Figure 60.1

Sounds approaching listener from two directions.
Parallelograms show cross-sectional areas for the two sounds.

Chapter 61. Energy Spectra

61a. Vibration Energy Spectra

The chapters about power and intensity following Chapter 49 sidestep the oscillatory nature of sound. The conservation of energy principle is so powerful that many questions can be answered on that basis alone. But this chapter combines energy ideas with the spectrum concept, which means getting frequency involved. Even without knowing details about what is physically oscillating, the energy in a complex vibration or sound can be conceptually divided between various frequencies.

In Chapter 42, Fourier analysis leads to the idea of a spectrum, which represents a complex vibration by breaking it up into component sinusoidal partials, each at a specific frequency. Each partial has a certain size, which in Chapter 42 is measured by an amplitude. But another way to describe the size of a partial is by the energy associated with it. Whichever energy-related concept is appropriate to the general situation, that same concept would be used for partials of the spectrum. For vibration of a system, we would use the energy contained in the vibration. In the case of a sound, or any disturbance which moves from place to place, we would instead focus on the rate of energy transmission (that is, power) or on the intensity. Whether the situation calls for an **energy spectrum,** a **power spectrum,** or an **intensity spectrum,** they share many characteristics.

Figure 61.1 shows an example. Part (a) is an amplitude spectrum (the same spectrum as in Figure 42.1), and part (b) is the **energy spectrum** for the same vibration. The two spectra have the same number of partials, which are at the same frequencies. This consistency is critical to understand before addressing the differences. The complex vibration is made of the same partials at the same frequencies, regardless of how we choose to measure their strength.

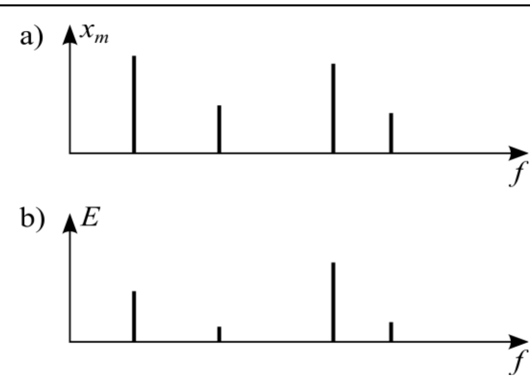


Figure 61.1

(a) Amplitude spectrum. (b) Energy spectrum for same sound

But there is a change: the heights of the peaks show a somewhat different pattern. The two spectrum types even show a different partial to be the "largest." What has happened?

Section 32b describes a close, generally applicable relationship between energy and amplitude. However, the energy associated with a vibration can also depend on frequency; the exact mathematical relationship depends on the nature of the vibrator, and even depends on what type of amplitude is specified. Thus, complications arise when dealing with multiple frequencies, such as in a spectrum. There is no universal rule for dealing with those complications.

This begs the question, when can we, or can't we, use proportion 32.1, the relationship between amplitudes and energy-related quantities?

- Proportion 32.1 safely applies to changes in the size of a vibration or sound as a whole. If a vibration or sound is uniformly scaled larger or smaller, so that all the partials change by the same multiplier, then the proportion can compare any amplitude to any energy, before and after the change.
- Proportion 32.1 also applies at each frequency separately. If a single partial changes size, the proportion can relate that partial's amplitude to its energy, regardless of what is happening to the other partials.
- Proportion 32.1 sometimes *cannot* be used to compare two partials at different frequencies, or for any change to a vibration or sound that involves changes of frequency.

Now that there is a choice for what to use on the vertical axis of a spectrum, which choice is best? Displacement amplitude is often the easiest to see and measure. Amplitude also relates most directly to superposition. But power and intensity more directly relate to how loud the partials are. Energy-related spectra also allow for the following rule arising from energy conservation.

> The total energy in a complex vibration is equal to the sum of the energies of the partials in its spectrum.

In contrast, Chapter 43 points out that summing the amplitudes of spectrum peaks yields a result with no practical meaning. The reason that this should work for energy is perhaps best illustrated in the case of sound. Given a complex intensity spectrum for a sound that you hear, we can imagine that the sound actually comes from multiple separate sources, one source for each of the partials. This is possible because spectra combine by simply merging their lists of partials. However, energy conservation tells us that the total intensity is equal to the sum of the intensities from each of the sources. This scenario can only make sense if the above rule works.

61b. Extra: Energy and Amplitude in Spectra

This section explains more details about why the two spectra in Figure 61.1 look different, even though they represent the same sound.

When a pure tone causes a sympathetic vibrator to respond with a sinusoidal motion, this is not a natural oscillation of the vibrator because it's not at the vibrator's natural frequency. As Chapter 32 explains, the energy in the sympathetic vibrator is not constant. But the energy does vary periodically with each cycle of motion, so there is an average value for the energy. That *average* energy is what is actually plotted on the vertical axis of Figure 61.1(b). When a sound causes a complex vibration in a sympathetic vibrator, even though the motions from the various partials are intermingled by superposition, it turns out that an average energy can be clearly defined for each partial.

Figure 61.1 illustrates the results in the specific case of a sympathetic vibrator with a specific mass and restoring force, and thus with a specific natural frequency f_0. Sorting this out involves math a bit beyond the level of this book. But the end result is that the average energy in a partial depends on both the amplitude and the frequency of the partial according to the equation

$$E_{\text{avg}} = \tfrac{1}{2}k\, x_m^2 \;\times\; \tfrac{1}{2}\left(1 + \left(\frac{f}{f_0}\right)^2\right) \quad . \tag{61.1}$$

That is, the average energy is the usual maximum potential energy, multiplied by a factor that depends on the frequency. Notice that if $f = f_0$, meaning a natural vibration, then the multiplier reduces to one, so that the natural motion result, as described in Chapter 31, still applies. The result of this relationship is that in the energy spectrum higher-frequency partials are more emphasized than lower-frequency partials.

For the purposes of this book, the exact form of Eq. 61.1 is not important. It serves only as a specific example of what it means for the energy in a partial to depend on both amplitude and frequency.

Chapter 62. SIL Spectra

If you have read Chapter 93, you will notice that this subject very much parallels that chapter.

Considering Chapters 57–59, it will perhaps come as no surprise that the size of the partials in a sound intensity spectrum can also be measured in SIL. The equation relating the intensity of a partial and the SIL of that partial,

$$L_I = (10\text{ dB})\log\left(\frac{I}{I_0}\right) \quad , \tag{62.1}$$

is exactly the same as the equation for sounds in general. Intensity and sound intensity level can be directly calculated from each other. Figure 62.1 shows an example of the same sound represented in both ways. A few features bear comment.

A relatively wide range of intensities is compressed to a relatively small range of SILs. This is one key aspect of level measurements (that is, measurements in dB). It's exactly the reason that SIL is helpful for describing the wide range of audible intensities. As a consequence, if the frequency axis were positioned to cross the origin of the L_I axis in Figure 62.1(b), the vertical sizes of the four partials in that graph would be almost indistinguishable. For level measurements, it is usually a good idea to choose a range for the L_I axis that is not much larger than the range of peaks to be included. This means that the frequency axis will usually not cross at 0 dB; 0 dB won't even appear on the graph.

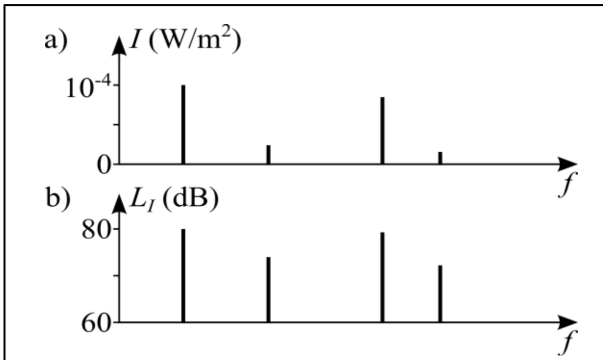


Figure 62.1
(a) Intensity spectrum. (b) SIL spectrum for same sound.

Although Eq. 62.1 is available to relate corresponding peaks (that is, an intensity peak and an SIL peak at the same frequency, both representing the same partial), one can also calculate peak heights by using the relationships between partials at different frequencies. As a specific example, suppose that we focus on the relationship between the first two partials in Figure 62.1. Eq. 58.2 gives the relationship

$$L_{I2} - L_{I1} = (10\text{ dB})\log\left(\frac{I_2}{I_1}\right) \quad . \tag{62.2}$$

Just as with sounds in general, the ratio of intensities of partials gives the difference of the SILs of those partials. In our specific example, the intensity ratio is

$$\frac{I_2}{I_1} = \frac{0.25 \times 10^{-4}}{1 \times 10^{-4}} = 0.25 \quad , \tag{62.3}$$

which directly gives the SIL difference in Figure 62.1(b)

$$L_{I2} - L_{I1} = (10\text{ dB})\log(0.25) = -6.02\text{ dB} \quad . \tag{62.4}$$

Depending on what information you have available, this method may be more convenient than calculating each SIL from the corresponding intensity. Or it can serve as an alternate perspective to check a direct calculation.