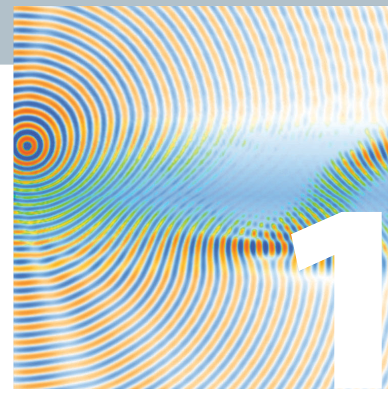


How Sound Propagates

Sound takes place when bodies strike the air, … by its being moved in a corresponding manner; the air being contracted and expanded and overtaken, and again struck by the impulses of the breath and the strings, for when air falls upon and strikes the air which is next to it, the air is carried forward with an impetus, and that which is contiguous to the first is carried onward; so that the same voice spreads every way as far as the motion of the air takes place.

—Aristotle (384–322 BCE), *Treatise on Sound and Hearing*

More than two thousand years ago, Aristotle correctly declared that sound consists of the propagation of air pressure variations.

Even to the casual observer, sound is plainly revealed to be a consequence of vibrating or pulsating objects in contact with air. Surfaces feel a force from all the molecules colliding with them; every molecule bouncing off the surface gives it a tiny shove. A bounce is a change of velocity and thus an *acceleration*, imparting a force F on the molecule (and an equal and opposite force acting on the surface) according to Sir Isaac Newton's law $F = ma$, where m is the mass of the accelerated molecule, and a is the acceleration.

Fluctuations of pressure above (*condensations*) and below (*rarefactions*) the average pressure, arriving at the surface as sound, cause a very small increase or decrease in the number of collisions per second, and a corresponding tiny but measurable change of force on the surface. These fluctuations above and below the ambient pressure are called the *pressure amplitude* δP, where the total pressure is $P = P_0 + \delta P$, and P_0 is the ambient pressure. Usually only the amplitudes matter to us; it is changes in pressure that we hear, not the ambient pressure. We (and other animals) however are *spectacularly* sensitive to these changes; a pressure fluctuation of just a few parts in a billion (a few billionths of an atmosphere) is enough for us to hear if it happens fast enough.

As small as it is, the tympanum is huge on the molecular scale. There are so many molecules colliding with it every millisecond (roughly 10^{23}—that's 1 followed by 23 zeros) that they average out and give a *nearly* steady pressure, amounting to about 14 lb of force on every square inch. Air pressure is usually measured in kilopascals (1 kPa = 0.145 pounds per square inch, or psi). Sea-level air pressure is about 100 kPa, or 14.5 psi. The tympanum membrane, which separates the middle and outer ear, normally has equal air pressure on both sides, so there is no net force on it, except for tiny fluctuations.

Aristotle could not have known that air is a seething mass of molecules crashing into one another. More than a billion collisions are suffered by every molecule every second at sea level and room temperature. In spite of all the collisions, air is mostly empty space: the molecules occupy only about one part in 5000 of the available volume. Think of 10 bumper cars in an area the size of a football field. You might think that this was a relatively safe, low density of cars—unless each car was traveling at thousands of kilometers per hour. There would be many collisions every second. Between collisions, molecules speed along a straight path at typically half a kilometer per second, managing to travel only a tenth of the length of a typical bacterium before suffering another collision.

The density and speed of air molecules are in this way sufficient to explain atmospheric air pressure and the speed of sound. Individually, the air molecules (mostly diatomic nitrogen and oxygen) act like drunken messengers flying and colliding every which way. Nonetheless, these collisions can collectively communicate even slight fluctuations in pressure to neighboring collections of molecules, which in turn pass them on to their neighbors, leading to sound propagation. Air molecules are usually not traveling directly along the path of the sound wave; the information that there is higher or lower pressure somewhere propagates no faster than the average speed of molecules along a given direction.

The typical 500 meter/second (m/s) molecule is traveling either in the wrong direction or only 300 to 400 m/s along the direction of propagation of the sound. Thus the effective speed with which the morass of molecules communicates pressure variations is less than their average speed of 500 m/s. The measured speed of sound in air is about 343 m/s at room temperature.

The "seething mass of molecules" picture explains why the speed of sound is insensitive to pressure, since pressure hardly affects the speed of individual molecules. They crash into each other more often at high pressure, but between collisions they travel at a speed that depends only on the temperature, not the pressure. The speed of sound on Mount Everest is nearly the same as at sea level, if the temperatures are the same.

The average speed of molecules is proportional to the square root of the temperature, and inversely proportional to the square root of the mass of the molecules in the gas.

"Helium voice," the Donald Duck–like sound when someone speaking has just inhaled a puff of helium, is the result of the much higher speed of sound in helium than in air. Helium has a mass of four atomic units; air has an average mass of about 29 atomic units and $\sqrt{29/4} \approx 2.7$. The speed of sound in helium, 972 m/s, is about 2.8 times that of air, at 343 m/s. Another harmless gas (except that like helium, it displaces oxygen and can be lethal if breathed for more than a short time), sulfur hexaflouride, SF_6, is much heavier at 146 atomic units and should have a speed of $343 \times \sqrt{29/146} = 153$ m/s; the measured value is 150, less than half the speed of sound in air. "SF_6 voice" is even more astonishing in its effect than helium voice, and in the opposite direction. (However, the nature of and reasons for the changes in the sound of the voice using helium and SF_6 will be explained in section 17.9. In spite of impressions, the gases do not change the pitch of the voice!)

The energy needed to make audible sound is very small. You can shout for a year, and the energy produced *that winds up as sound* would not be enough to boil a cup of water. A full orchestra playing loudly produces only about enough sound energy to power a weak lightbulb. An orchestral crescendo might bathe a listener in sound pressure fluctuations of about 1 pascal (1 Pa). Sea-level air pressure is 100,000 Pa, so the crescendo loud enough to damage your hearing, if it lingered too long, is varying the pressure by just 0.001%. Clearly, a very delicate detection system is at work. We will find in chapter 21 that human hearing depends on a few thousand *single-molecule* links between cochlear hair cells.

At the extreme—loud sound near the threshold of pain—the air pressure variations are over a million times bigger than the threshold of hearing, or about a 0.03% pressure variation, 30 Pa or so. This still seems small, and yet is almost immediately damaging! This sound level corresponds to a power arriving at the ear 10,000,000,000,000 (10^{13}) times larger than that which produces the softest sound we can hear. (The power increases as the square of the pressure variations.) The dynamic range of our hearing is truly remarkable.

Why should you buy a 600-watt (W) amplifier for your loudspeakers if a full orchestra normally produces just a watt of power, 40 or 50 W at the loudest? The answer is that to reproduce sound, rather large forces must be exerted on the speaker cone to get it to vibrate in a prescribed way. The conversion efficiency from motion of a loudspeaker cone to sound is very low. The cones are moved with electric currents in coils near magnets, wasting considerable energy. Imagine all the effort you would expend waving your hands back and forth 1000 times. Only a tiny fraction of that energy would go into pushing air around; most of the energy expended would go into working against yourself, so to speak: internal friction, stopping your arms with one set of muscles after starting them swinging with another, working against gravity, and so on. So it is with a loudspeaker. For that matter, musicians can work up a sweat playing an instrument, all to produce well under a tenth of a watt of sound.

1.1 Push and Pushback: Impedance

We need to develop a better intuitive foundation for sound propagation. The "drunken messenger" picture explains the speed of sound but applies on the molecular scale, too small to give a good sense of wave phenomena such as reflection, diffraction, refraction, and so on. For example, much of the sound traveling down a tube reflects from its open end, reversing direction rather than exiting to freedom. Why doesn't the sound just leave the tube? Why is the reflected wave a *rarefaction* (pressure low relative to ambient) if the incident wave approaching the end of the tube was a *compression* (pressure high relative to ambient)? Why does sound of high-enough *frequency* (the frequency is the number of wave crests traveling by per second), on the other hand, mostly escape the tube without reflecting? There are not many references that provide a foundation for a comprehensive understanding of these sorts of phenomena; those written for engineers and physicists all too often derive equations and formulas but skimp on the intuition.

Imagine dividing air into small cells. Each cell is large on the molecular scale; they are packed one next to the other. The size of the cell is determined by the wavelength of the sound involved (there needs to be at least several cells per wavelength) and the details of any obstacles, sound sources, and so on. If we can understand how the cells communicate with each other, are pushed by and then push back on neighboring cells, we can understand propagation, reflection, diffraction, and even refraction of sound. This is our first glance at a powerful engineer's trick, wherein the properties of complex objects are lumped into a few well-chosen summarizing properties. These have vastly less information than the original system, but enough to carry the essential physics, and lead more easily to the correct conclusions.

To understand impedance in air, we begin by considering solid elastic bodies, such as pucks on an air hockey table or coins on a slick surface. We need to understand such things in any case, because usually before air is set in motion, something more massive, like a string or a block of wood, is set in motion. Each puck or coin is a lumped object—we ignore the details of atomic or molecular structure inside, but keep essentials such as size, density, and elasticity, just as we will for air cells when we return to them. The essentials are used to build a theory of what happens when adjacent lumps interact. You may have noticed, for example, that in a head-on collision between two pennies, one initially at rest, the moving penny stops dead in its tracks, and the second one picks up where the first one left off. (The demonstration does not work well with quarters or coins having serrated, gearlike edges colliding with other such coins. Presumably the serrations cause a rather nonideal collision, gnashing of the gears, chattering, and so on.)

Complete transfer of momentum does not happen when a nickel collides with a penny at rest, nor when a penny collides with a nickel at rest. The energy of the first, moving coin is only partially transferred to the second. If we make a line of coins, each coin becomes an agent of transfer of energy from left to right, if the first coin was traveling in that direction. A coin that is much heavier or lighter than its neighbors will impede the transfer of energy. Two nickels are *impedance matched*; the stationary nickel gives as good as it gets, stopping the moving nickel dead. A penny and a nickel are impedance mismatched; a penny does not exert as much force back on the colliding nickel as another nickel would and does not decelerate the nickel all the way to zero speed. The nickel continues on its way, albeit more slowly. Only in the case of equal masses does the energy get completely transferred from one coin to the other; this is clear since for the head-on nickel-penny and penny-nickel collisions, both coins remain in motion and that movement carries energy.

If you line up 5 or 6 pennies perfectly on a slick surface and hit the end of the row head-on with another penny, you will notice the row stays intact, with the projectile penny adding to the row and the last penny popping off at the opposite end. The impedance matching works all the way down the penny chain, each penny for an instant carrying the momentum, giving as good as it got on its left, and then almost instantly giving and getting forces on its right that stop it cold and give the momentum to the next penny. Put a nickel in the chain of pennies and the first penny will rebound from the row; the last will still pop off the end but with less energy than before. All the energy of the first coin is not transferred down the chain; rather, part of the energy has been reflected and part transmitted, *because of the impedance mismatch, which can be blamed on the interloping nickel*. The situation is depicted in figure 1.1.

The impedance of the untethered coins is proportional to their mass. Two untethered objects of equal mass, therefore, indeed have the same impedance. The bigger the impedance mismatch, the more energy is reflected and the less transmitted. The formula for the fraction of energy

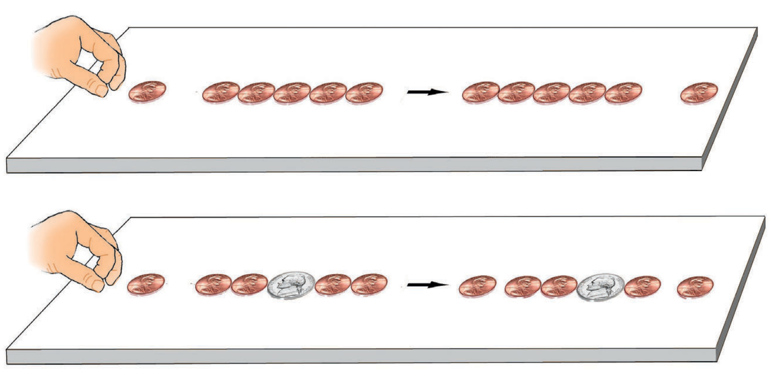


Figure 1.1

In the top row, a penny collides head-on with a row of five pennies, resulting in the expulsion of the last penny in the row with the same speed as the first penny had. The masses are all the same and the chain of pennies is impedance matched, resulting in 100% transfer of the energy from the first penny to the last, except for friction. In the bottom row, the presence of the nickel replacing one of the pennies causes a mismatched impedance, with some of the energy reflected back toward the first penny, causing it to rebound; only part of the energy reaches the last penny.

R that the moving mass m_1 retains in a head-on collision with a stationary mass m_2 is

$$R = \frac{(m_1 - m_2)^2}{(m_1 + m_2)^2}. \quad (1.1)$$

If one coin weighs one-tenth as much as the first, say, $m_1 = 1$, $m_2 = 10$, R will be $9^2/11^2 = 81/121$, which means 67% of the energy gets reflected on one bounce, and 33% transmitted.

What Is Impedance, Really?

Roughly speaking, impedance, which we symbolize with the letter Z, measures the response of a body to a force—in fact, the force applied divided by the velocity attained ($Z \sim$ force $\div$ velocity). A heavy object moves slower than a light one after the same force is applied starting at rest, so impedance is high for a heavy object, low for a light one. This is still a rough definition, since in the measurement of Z, the force is taken to vary sinusoidally (see chapter 3), and the velocity, while also sinusoidal, may lag or lead the force. We will consider these complications later.

With this notion of Z (force applied $\div$ velocity attained), it is possible to see why matched impedance leads to complete energy transfer between two bodies. According to one of Newton's laws, they experience equal and opposite force as they collide or interact, and what velocity is lost by one is gained by the other—just the ticket if you want to transfer energy from one place to another, or from one thing to another. One coin stops and the other takes off with the same velocity.

The utility of impedance is to help determine the transfer of energy between bodies. Matched impedance means efficient energy transfer; unequal impedances mean rejection or reflection of energy. Ideally, impedance can be determined for any part of an object, such as a block of metal or a section of pipe with air in it. If two such objects are joined somehow, an impedance mismatch (if any) can be calculated, and the *transduction* (transfer) of energy from one part to another can then be determined.

As an example, suppose two strings of different density are tied together. We will see in chapter 8 that waves travel down a uniform string quite readily, with a velocity $c = \sqrt{T/\rho}$, where T is the *tension* (a force) along the string, and ρ is the *density* (mass per unit length) of the string. The two parts tied together have the same tension, since tension is communicated all along the string, but they have different density, and thus different wave speeds c. They also have different impedances. The impedance of transverse oscillations of a stretched string is

$$Z = \sqrt{T\rho}. \quad (1.2)$$

Given the densities ρ_1 and ρ_2 of the two string segments, we can easily calculate the reflection and transmission of energy at their junction using formulas 1.3 and 1.4 given below.

Antireflection Strategies

Suppose we insert a third coin between two mismatched coins, one more massive than the other. The middle coin should be of some intermediate mass, to make the mismatches of adjacent coins less severe. It is not difficult to show that taking the mass of the middle coin to be the geometric mean of the two original coins (that is, $m = \sqrt{m_1 m_2}$) is optimal. The transmission with the intermediate coin in place in the 1:10 impedance mismatch considered earlier then works out to 53% from the first to the last coin; an improvement over the previous 33%. We would do even better with more intermediate coins selected to further reduce the adjacent impedance mismatches.

Abrupt changes in impedance at a boundary between two objects or regions lead to low transmission of energy across the boundary. Like the nickel in a line of pennies, regions with different impedance push back too much or too little. Suppose we have a system of one impedance Z_1 on the left side connected to a second system on the right with a different impedance Z_2. The sudden change of impedance at the interface causes a fraction of energy $\mathbf{R}$ to be reflected:

$$\mathbf{R} = \frac{(Z_1 - Z_2)^2}{(Z_1 + Z_2)^2}. \tag{1.3}$$

Thus equation 1.1 generalizes to more general types of impedance, including (as we shall see) restoring force and friction. The *transmitted* energy is

$$\mathbf{T} = 4\frac{(Z_1 Z_2)}{(Z_1 + Z_2)^2}, \tag{1.4}$$

and the reflected and transmitted fractions sum to one: $\mathbf{R} + \mathbf{T} = 1$—that is, what is not transmitted is reflected.[1]

Impedance matching plays a role in many domains. In the preceding example, the coins were a "medium" for the propagation of the translational energy possessed by the first coin. Light is similar: it propagates nicely through transparent media, such as air and glass, but these do not have the same impedance. The impedance (called *refractive index* in the case of light) has a mismatch passing from air to glass, with the result that some

[1] The impedances are in fact complex numbers, so we have $\mathbf{R} = (|Z_1 - Z_2|^2/|Z_1 + Z_2|^2)$ and $\mathbf{T} = 4[\mathrm{Re}(Z_1^* Z_2)/|Z_1 + Z_2|^2]$, where Re denotes the real part of the variables within the parenthesis, and $|\cdots|^2$ is the absolute value squared of $\cdots$.

Figure 1.2

What would this sound like? A string is attached directly to a violin body at one spot (no bridge) and to a rigid wall at another. It is bowed in the usual way.

light will reflect at the interface, whether it is coming from air to glass or vice versa. If a coating can be found with intermediate impedance, it can break up the impedance mismatch into two smaller steps, with the result that less light will be reflected and more transmitted. This is the principle of antireflection-coated eyeglasses and camera lenses. The coating works better for some colors (wavelengths) than others; this explains the color sheen often seen on coated optics.

As an example of the importance of impedance to sound and music, consider a violin. The body of a violin is much heavier and stiffer than a string and has a much higher impedance. Both impedances vary with frequency too. The body needs to tap into the energy of the string in order to make sound. (Vibrating strings by themselves are almost silent—this will be made clear in the following chapters; see especially the discussion of dipole sources—for example, section 7.7). Hypothetically the string could be attached directly to the body, but there are several problems with this (see figure 1.2). The directly connected string may not set the correct body vibrations into play. Worse, there is a large impedance mismatch between string and body, preventing the string from imparting enough of its energy to the violin. (Note: We don't want the transfer of energy from string to body to be *too* efficient either, lest the string dump its energy too fast.)

Impedance and the Violin

Air has a refractive index n_{air} of about 1, and glass can be $n_{glass} = 1.5$ or so. The refractive index is essentially impedance; the formula for the fraction of light reflected is

$$R = \frac{(n_1 - n_2)^2}{(n_1 + n_2)^2}. \tag{1.5}$$

This is a 4% reflection of light for air–glass, for each surface, and there are always at least two surfaces and sometimes many more, as in expensive camera lenses. By adding an optimal single coating, with the geometric mean refractive index $\sqrt{n_{air} n_{glass}}$, we can get this down to a 2% reflection. Multiple coatings can do even better.

Can something be inserted between string and violin body to lessen the impedance mismatch, thus allowing the energy to take two smaller steps, instead of one large one? While we are at it, can we sweeten the sound by modulating the impedance (and ultimately the loudness of the instrument) according to frequency? The answer is yes: this is the job of the bridge, as we discuss in chapter 18. The bridge is the "intermediate coin" that mediates the transfer of energy from string to body. Its impedance is cleverly tuned by choice of shape, size, and material to depend in a certain way on the frequency of vibration.

Bullwhip—The High Art of Impedance Matching

The bullwhip is a spectacular example of impedance matching (figure 1.3). If most of the energy from the relatively heavy handle region can somehow be efficiently transferred to a light string ("popper") at the other end, the popper will wind up moving very fast. Sudden impedance mismatches along the whip would reflect energy, so the bullwhip is gradually tapered and also carefully constructed so as to have no abrupt changes in density or stiffness. The energy of a moving mass m due to its motion is $E = \frac{1}{2}mv^2$, where v is its velocity. A reasonable estimate is that the popper weighs 1/400th as much per centimeter of length as does the handle end. The energy per centimeter if the handle region weighs M kilograms per centimeter is $E = 1/2MV^2$, where M is the mass of a centimeter near the handle end, and V its initial velocity. If this gets transferred to the popper, then the same energy is now written $E = 1/2mv^2$, where m is the mass per centimeter of the popper, and v is the velocity of the popper. The ratio of the two velocities is

$$\frac{v}{V} = \sqrt{\frac{M}{m}} = 20 \tag{1.6}$$

in this case. A factor of 20 does not sound huge, until you realize it is easy to get the handle moving at 40 miles an hour (a fast baseball pitch is 100 miles per hour), and 20 times that is 800 miles per hour, or faster than the speed of sound at 770 miles per hour! The popper thus goes *supersonic* (faster than the speed of sound). A supersonic object traveling through the air creates a shock wave, a very sharp pressure pulse. (More on supersonics and shock waves in section 7.9.) The pulse itself travels through the air at the speed of sound, but when it reaches the ear, it is heard as a loud bang.

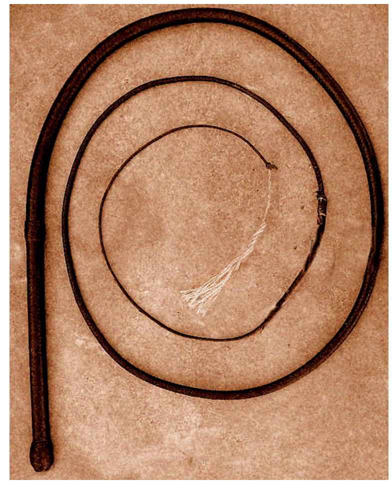


Figure 1.3

An Australian bullwhip can achieve supersonic speeds at the whip end, resulting in a loud crack heard some distance away. Courtesy Cgoodwin.

Impedance Mismatches Are Not Always Bad

One does not always want to maximize energy flow across junctions between two parts of a system. We *need* the impedance mismatch at the bell end of a trumpet or clarinet to cause reflections and define its resonance frequencies. Impedance mismatches are carefully controlled to achieve desirable timbre in wind and string instruments. For string instruments, large mismatches are required at the points between which strings are stretched, lest the vibrations drain away too rapidly, rendering the string frequencies ill-defined. The infamous wolf note of cellos is a breakdown of this requirement (see section 18.7)—a near impedance matching where none was wanted.

Impedance of Masses and Springs Together

Untethered coins on a table move along without hindrance (except for friction, which we have neglected so far) but many objects are tied down and experience a restoring force pulling them back if they are displaced. The concept of impedance applies, but now impedance can be high owing not only to large mass but also to large stiffness because of a spring, which also tends to keep speed low. A mass and spring can combine to make an oscillator that vibrates at a certain natural frequency; if you push back and forth at that frequency, the impedance is low even if the mass is large and the spring is strong, because the oscillator gets moving very fast.

Three universal properties of matter figure into impedance: (1) *Mass* is responsible for resistance to *acceleration*, as is encoded in Newton's second law of motion $F = ma$ (force = mass × acceleration). For a given force, acceleration and mass are inversely proportional. (2) *Stiffness* is responsible for resistance to being stretched or compressed, as encoded in the spring equation $F = -kx$, where F is the force, k is the spring constant, and x is the displacement. (3) The third universal property is *friction*. We are deferring that topic for the moment; see section 10.6.

If the force is applied slowly, acceleration is weak. The force is then usually governed by compressibility or springiness, which therefore gives *stiffness-dominated impedance*. If a force is applied suddenly, the object hardly has time to move and sense its stiffness, but the mass of the object is felt immediately; the impedance is mass dominated.

Defining and Measuring Impedance

We measure impedance by applying a back-and-forth, sinusoidal forcing. (The sinusoid is the subject of chapter 3.) The impedance will depend on the frequency of this forcing. If the *period* (time interval between repetition of the periodic force) of the forcing is very short (high frequency), then the force is changing suddenly; not much movement of the object takes place because such a short time elapses between reversals of the force. The impedance will tend to be mass dominated. If the frequency is low and the forcing period is very long, then the force is being applied slowly; the impedance will tend to be stiffness dominated. The object or matter in question is forced according to $F(t) = F \sin(2\pi f t)$; this periodically pushes right and left with frequency f. The sine function never gets bigger than 1, so the maximum force is F.

The object or matter being forced sinusoidally will temporarily build up speed in one direction and then slow down, stop, and reverse direction, building up speed in the opposite direction. Reaching high speed suggests a large response to the forcing, which in turn implies that the object presents low resistance—that is, low impedance, to energy *at the forcing frequency* f. The frequency-dependent impedance $Z(f)$ is defined as the ratio of the maximum force F to the maximum speed $u(f)$ reached at that frequency f:

$$Z(f) = \frac{F}{u(f)}. \tag{1.7}$$

The higher the speed $u(f)$ attained, the lower the impedance. This definition still ignores the phase lag or lead of the velocity relative to the force. The impedance $Z(f)$ used by engineers is a complex number—that is, it contains the imaginary number $\sqrt{-1}$; however, we will suppress that fact. (The information we throw away by doing this is the phase [see chapter 3] of the velocity attained relative to the forcing. We will discuss the phase quite thoroughly, but will not write it in terms of complex numbers.) Impedance is a measure of the ratio of the (sinusoidal) force applied to the speed attained. If we apply a large force and don't get much speed out of our efforts, the impedance is high. If for the same force, the point where the force is being applied reaches a high velocity, the impedance is low. It is important to remember that in our simplified version of impedance, the force is calculated as the maximum force at the point of application, and u is the maximum speed attained by that point.

To measure impedance, we can control the force and then measure the resulting speed—that is, control the numerator and measure the denominator in expression 1.7. Or we can control the speed of the point of application, and then measure the force that is needed to maintain that speed—that is, we can control the denominator and measure the numerator in equation 1.7. If the force or the velocity is controlled at the same spot on the object, the same value for the impedance is obtained either way. Extended objects will have different impedances depending on where the force is applied.

1.2 Impedance of Air

The idea of "push and pushback" and impedance can now be made more precise for air. Air has mass and is springy—so there ought to be a way to connect air to the impedance ideas we just discussed. Again we arbitrarily divide up a body of air into cells. The cell walls are purely mathematical—completely elastic and having no mass of their own. They do not exert any force or pressure of their own, but rather just follow along with the adjacent air. This division into cells does no harm, yet it helps our thinking. Each cell has mass and springiness. It is in contact with other cells with their own mass and springiness. Taking the cells to be cubic, if we push on one side of a cell, it will tend to bulge out on five other sides.

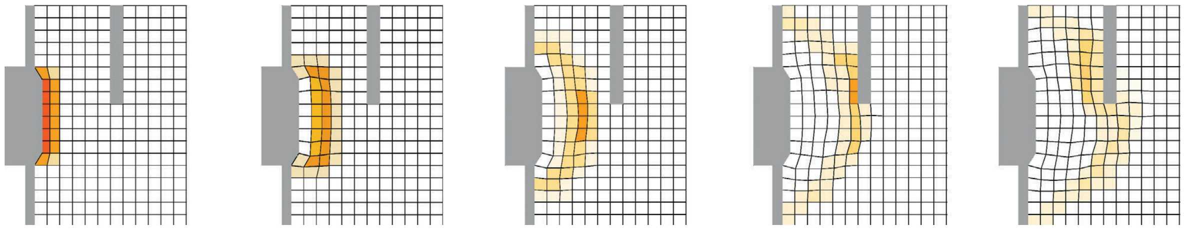


Figure 1.4

A piston initiates a pressure pulse in the cellular picture of sound propagation. Propagation, reflection, and diffraction are all represented.

The restoring force that any given cell presents upon being pushed on one side depends on how much pushback it receives when it tries to bulge out on the other sides. If one side of the cell is up against a rigid wall, the pushback from pushing on any other side will be higher, since one side can't move at all. Thus the presence of the wall causes an impedance change.

The impedance of a cell of air has three components: a component due to the mass of the air inside, a component due to the restoring force or springiness of the air, and a component due to friction, which we can safely ignore if the air is far enough from surfaces. In analogy with our line of pennies, cells of air are stacked next to each other, in three dimensions rather than one. Normally, each cell of air is just like the ones adjacent, which strongly suggests that air is impedance matched with itself and will efficiently transmit propagating sound.

Let's see how this works to explain the propagation of sound. Figure 1.4 shows a sequence of five snapshots in the evolution of a cell system with walls and a piston present. On the left, a piston has just pushed into the area, causing a region of high pressure next to its surface. Each cell contains the same quantity of air, so smaller cells are higher pressure. The piston holds its place, and the pressure wave begins propagating by the "shove and be shoved" principle. The color shows the pressure, and the distortions of the walls of the air cells are shown. A half-wall mid-chamber intersects the wave, and in the last frame we see reflection and diffraction from the wall well under-way. The cells just next to the piston are compressed initially, but they shove their neighbors and return to normal pressure. The domino effect continues as the wave propagates.

How big do the cells need to be? There is no single answer to this question, because a few smaller cells can often be replaced by one bigger cell, but there is a limit: the cells need to be much smaller than the shortest important sound wavelength, so that the information that they are being pushed on one side travels to the other sides in a time much shorter than a period of the sound. Usually a few centimeters or, at worst, a few millimeters on a side (giant on the scale of the distance between atoms and molecules in air) will suffice. In free space, they can be about a tenth of the smallest wavelength present, or even larger. But there may be solid objects or density changes on a much smaller scale than the wavelength, which rudely interrupt the wave. If their effect is to be included accurately, especially if the listener is nearby, smaller cells need to be used near such objects.

If a cell pushes back too hard (higher impedance than its neighbor), then the neighbor doing the pushing will recoil, pushing back on *its* neighbor on the opposite side, causing a positive pressure pulse to propagate backward—a *reflection.* If the adjacent cell, on the other hand, pushes back too feebly (lower impedance than its neighbor), then the pushing neighbor will keep moving toward the weak neighbor, ultimately *pulling* on *its* neighbor on the opposite side. That neighbor pulls in turn on *its* neighbor on the opposite side, and so on. A *rarefaction* is propagating back toward the source. A positive pressure fluctuation will thus partially reflect back as a negative one if it meets reduced impedance. If the adjacent cells are impedance matched, each pushes back just enough so as not to reflect any of the pulse.

The impedance of water is about 3400 times larger than the impedance of air. You may have noticed that if you are underwater, it is very difficult to hear someone above water, even if he is shouting. Using formula 1.3 for the amount of energy reflected, we find that about 99.9% of the sound arriving from the air is reflected from the water surface. Sound launched within water travels quite well; if it reaches the surface, it reflects back down. Notice from formula 1.3 that the percent of energy reflected is the same, no matter which side of the interface the energy is approaching from.

Several types of impedances are used for air, depending on the situation. All of them are a ratio of a force to a velocity or, if you like, the ratio of a "push" to "flow."

Specific acoustical impedance z. The push or force is measured in fluids as pressure p—that is, force per unit area on a surface. The flow v is just the speed with which the small cell moves due to the pressure. The specific acoustical impedance is just the ratio of these two quantities:

$$z = \frac{p}{v}.$$

Again, we are glossing over the relative phase lag of the pressure versus the velocity; they may reach their maxima at different times under sinusoidal pressure variations.

If there are no surfaces or reflections of any sort, the specific acoustical impedance is an intrinsic property of the medium, given by the product of the density of the medium ρ_0 and the speed of sound in it c; $z = \rho_0 c$.

Acoustical impedance (lumped) Z. The specific acoustical impedance is determined at a single point. Sometimes a lumped impedance is better to work with. For example, when we want to determine the impedance mismatch and reflection upon a sudden change of pipe diameter, it is convenient to have a single lumped impedance for pipes of given diameter.

For this, the impedance definition is changed a little, so that all the cells across the pipe are lumped together and the velocity used is the *volume velocity*—that is, the velocity attained by the little cells times the area S of the pipe. For a pipe where the diameter is small compared to a wavelength, the velocity v as a sinusoidal wave passes by will be essentially uniform across the pipe, so the volume velocity is $U = v \times S$ and the acoustic impedance of the pipe of cross-sectional area S is

$$Z = \frac{p}{U} = \frac{\rho_0 c}{S}.$$

Thus the impedance of a pipe is inversely proportional to the area of the pipe.

In developing our "push and pushback" intuition for sound propagation, we are in fact coming very close to the way numerical computations are done. We will not go into the details of the algorithms here, but it is not difficult to imagine that a computer can be programmed to determine the result of all the pushing and shoving by air cells, including the effects of boundaries.

Keeping track of the air pressure variations everywhere, including the effect of various nearby surfaces, is an enormous task, even for twenty-first-century computers. However, by employing banks of *graphics-processing chips* (the computers within the computer that control screen display, called graphics-processing units, or GPUs), we can carry out the calculations required to simulate the generation and propagation of sound. GPUs became powerful and cheap primarily because of the demands of computer games. It will not be long before acoustical consulting firms will be providing their clients with accurate and perfectly detailed computer simulations of the sound in concert halls or other soundspaces, including the effects of curtains, statues, chairs, and people; sound absorbing surfaces of all sorts; open windows; and so on. The process of computing the sound pressure field—by following the movement of the sources of sound, the propagation of sound waves, and all the reflections, refractions, absorption, and so on that are present, turning it finally into a playable sound file—is called *auralization*.

1.3 Propagation of Sound in Pipes

Pipes make the whole issue of sound propagation much simpler, provided we confine ourselves to sounds whose wavelengths are long compared to the diameter of the pipe. Such long wavelengths propagate along the axis of the pipe but don't vary much from center to edge of the pipe, permitting a one-dimensional treatment in terms of the distance down the axis of the pipe. Pressure is given as a function of this distance and time along the pipe axis. This is much simpler than trying to work out all the variations in a three-dimensional sound field.

We suppose that such a pressure wave is traveling down the pipe. This is easily arranged in a number of ways, such as slapping an open end with a flat object. The propagation of such a pulse down a straight-walled tube is intuitive from the cellularization and impedance picture of the air in the pipe. First, since we will not be concerned with variations in pressure across the pipe, we can enlarge the cells into thin lozenges that extend across the pipe, taking on the cross section of the pipe. The pressure is taken to be constant everywhere in a given lozenge.

Each lozenge has mass m and is pushing out on its two neighbors; they push back just as hard in the quiescent state. If a disturbance arrives, a lozenge momentarily pushes on its neighbor a little harder, which now feels an unbalanced net force F as it begins to accelerate according to $F = ma$. The acceleration in turn induces a harder push on the next lozenge, and so on down the line, leading to propagation of the pulse.

We now turn to what happens when changes in the pipe are encountered by the pulse.

Reflection of Sound at a Closed End

The impedance of all the lozenge cells is the same because they are all identical in a pipe that does not change cross section. Suppose, however, the pulse meets a rigid end cap—that is, infinite impedance. The cell next to the wall pushes back on the adjacent cell very hard, since it has nowhere to go. This "over-pushback" causes the adjacent cell to recoil in the reverse direction; in turn, it pushes on its neighbor on the side away from the wall, and so on. There is thus a traveling pressure pulse that has reversed direction; it has bounced or reflected off the end cap with no loss of energy (figure 1.5). *Note that the end cap did not move at all to cause this reflection, or* echo*, of the sound.*

Reflection of Sound at an Open End

If a pipe terminates in an open end, it is much the same as a sudden very large increase in pipe diameter. We expect a sharp drop in impedance; the discontinuity will reflect sound amplitude back with the opposite sign. The

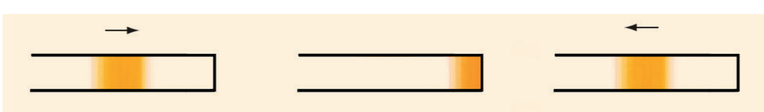


Figure 1.5
Reflection at a closed end cap in a pipe, taken directly from a Paul Falstad *Ripple* simulation. The simulation of a single half-wave, as seen here, can be set up in *Ripple* by initiating sinusoidal waves to the left of the pipes and later erasing all but one half-pulse inside the pipe before it reaches the junction.

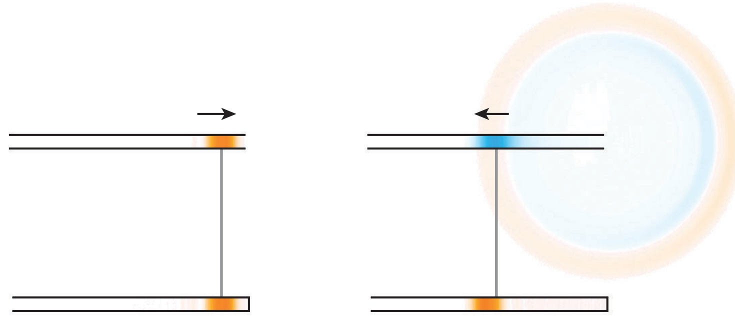


Figure 1.6
Reflection of a pressure pulse at the open end of a narrow pipe (top) and the closed end of a narrow pipe (bottom). Three significant effects are seen: First, the sign of the pulse reverses in the case of the open end, but not in the case of the closed end. Second, in the case of the open end, not much of the sound is emitted; most reflects. Third, there is a slight delay (as seen using the vertical reference line) of a pulse in the case of the open-end pipe as compared to the closed-end pipe, as if the open pipe were slightly longer. The delay is evidence of the *end correction* which makes open pipes effectively somewhat longer than their nominal physical dimensions.

air at the end of the pipe feels less pushback, overshoots, and pulls on the air behind it, initiating a rarefraction that propagates backwards.

Figures 1.6 and 1.7 show this effect quite nicely. Both figures are taken directly from *Ripple* simulations, which we can set up by drawing the pipes and sending in sinusoidal waves. The simulation is stopped, and the Erase Wave tool is used to trim the wave to lie inside the pipe and to be only half a wavelength across.

An open pipe partly reflects the wave with a change of sign. It reflects as if from a place just outside the end of the pipe, making the pipe effectively longer by about 0.6 times the diameter, for wavelengths that are large compared to the diameter.

As an interesting test of our understanding, suppose we send a pulse through a tube heading toward an open end, but this time the pulse exists outside the pipe as well. What will happen when the pulse reaches the end of the tube? The air inside the pipe has no idea that the pressure pulse exists outside until it reaches the end; as the pressure exits the pipe, instead of finding lower pressure laterally as it did before, it now finds matched higher pressure outside. There is no sudden pressure release laterally, no impedance change. The entire pulse proceeds as if nothing happened; there is no back reflection inside the pipe at all. Figure 1.7 comprises two snapshots from a *Ripple* simulation verifying this effect.

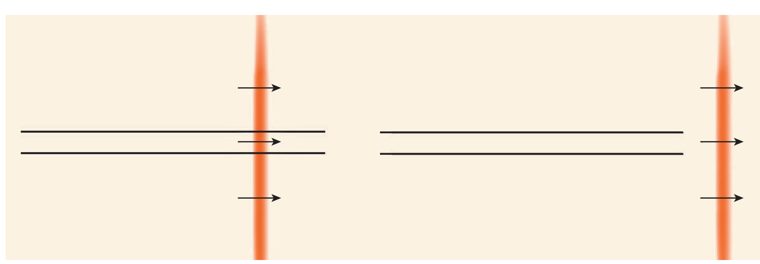


Figure 1.7
Two snapshots of a *Ripple* simulation showing a pulse propagating from left to right both inside and outside a tube. When the pulse exits the pipe, no reflection takes place.

Reflection of Sound at the Junction of Different-diameter Pipes

If the pipe changes diameter, the pulse will meet a change in impedance. Cells on the other side of the diameter change will push back too hard (if the impedance it meets is higher than its own), or too little (if the impedance it meets is lower). This will cause partial reflections of the sound at such junctions.

Earlier, we said that the impedance of air in a pipe depends on the diameter of the pipe. The bigger diameter, the lower the impedance. This makes a certain amount of sense, since a small pipe "impedes" the flow of air more than a large pipe. The impedance is again $Z_{pipe} = \frac{\rho_0 c}{S}$, where ρ_0 is the density of air, c is the speed of sound, and S is the cross-sectional area of the tube.

The physical reason for the increase of specific impedance as the pipe diameter decreases is understandable from the cellular picture. The higher specific impedance of a small pipe implies that if a small cell of air is pushed, a neighboring cell will push back harder than it would in a larger pipe. Why should this be? All the pushing and pushing back is of course communicated by the air in the pipe from cell to cell at the speed of sound. Suppose a given cell is being pushed to the right for a time τ; in free space, that push would be communicated in all directions a distance $x = c\tau$ in the time τ, where c is the velocity of sound. In the pipe, most of those directions lead to the walls of the pipe, where the pressure pulse created by the push is reflected. Some of the reflected pressure returns fast enough to the cell that was originally disturbed that it leads to an increased pushback, *while the original push is still happening and therefore in phase with the pushing*, thus increasing the impedance. "Fast enough" is in relation to the frequency of pushing. This suggests the wall needs to be within an eighth of a wavelength or so, to return in phase. Most musical instruments are

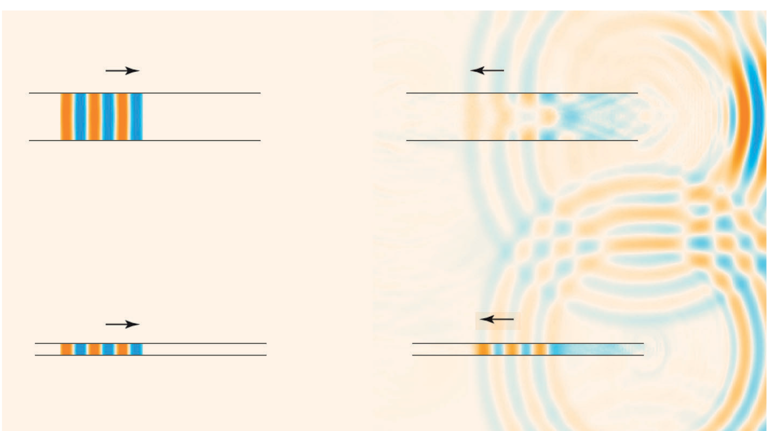


Figure 1.8

Sound of the same wavelength propagates in a narrow and a wide pipe in this *Ripple* simulation. It escapes more readily from the wide pipe, which can be seen by inspecting the intensity of the reflected waves in the right pair of panels. This can be justified using the cellular picture and impedance arguments, as explained in the text.

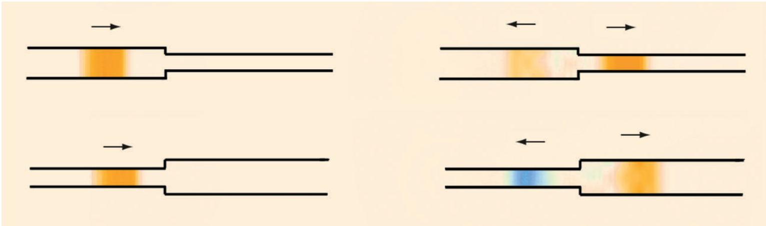


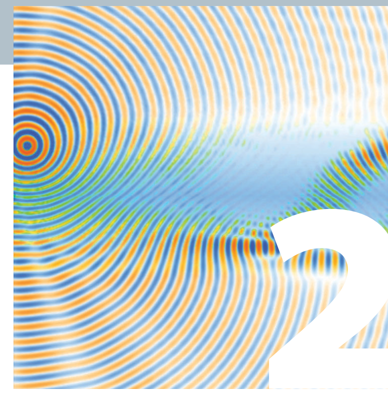
Figure 1.9

Reflection at a discontinuity in pipe diameter, taken directly from a Falstad *Ripple* simulation. (Top) A compression pulse traveling to the right encounters a smaller pipe, causing a compression reflection (same sign as the incident wave) and partial transmission of the compression pulse. (Bottom) A pulse of higher pressure (compression) traveling to the right encounters a larger pipe, causing the reflection of a rarefaction pulse (opposite sign from the incident pulse) and partial transmission of the compression pulse.

operating at frequencies such that the wall is always this close or closer. In fact, the pressure pulse doesn't reflect just once, but many times, depending on the diameter of the pipe. Thus the narrower the pipe, the higher the impedance.

The cellular picture confirms that short-wavelength sound will escape the pipe more readily than does long-wavelength sound. The frequency is higher for the shorter wavelength, so a cell just inside the pipe may not get an in-phase, reinforcing reflection from the walls in time to increase its impedance. It acts more like a free cell and thus doesn't notice much change as it encounters cells outside the pipe: little impedance mismatch, and little reflection. This is exactly what is seen in the *Ripple* simulation in figure 1.8, where a wave train of the same wavelength is traveling down a narrow and a wide pipe (right). After the encounter with the open end, much stronger reflection is seen inside the narrow pipe, and stronger transmission is seen outside the wide pipe (even accounting for the fact that there was more wave energy in the big pipe to begin with). Take note of the wavelength of the wave compared to the pipe diameter in both cases.

If a pipe suddenly becomes narrower, or wider, there is a corresponding abrupt impedance change (mismatch) at the junction of the two sections of pipe. If a positive pressure pulse is traveling from a wider to a narrower pipe, a positive pressure pulse returns from the junction, reflecting part of the energy. If instead it encounters a wider pipe, a negative pressure pulse reflects part of the energy (figure 1.9).



Wave Phenomenology

It happened once, on board a ship, sailing along the coast of Brazil, 100 miles from land, that the persons walking on deck, when passing a certain spot, heard most distinctly the sound of bells varying as in human rejoicing. All on board listened and were convinced, but the phenomenon was mysterious and inexplicable. Some months afterwards it was ascertained that the bells of the city of St. Salvador, on the Brazilian coast, had been ringing that very day on the occasion of a festival. The sound had, therefore, travelled over 100 miles of smooth water, and, striking the wide-spread sails of the ship, rendered concave by the breeze, had been brought to a focus, and rendered perceptible to all on board.

—Brewer, *Sound and Its Phenomena*, 1864, p. 288

Now is a good time to introduce some of the phenomenology of waves and their propagation. Developing an intuition for waves can be greatly aided by Paul Falstad's Java applet *Ripple*.

In figure 2.1, we show a simple sinusoidal wave, with its amplitude A and wavelength λ indicated. We also adhere to a convention, used throughout the book, that positive amplitudes are colored orange and negative ones, blue. This is, of course, unnecessary in a picture of a one-dimensional wave but is very helpful when plotting waves in two dimensions.

2.1 Relation between Speed, Frequency, and Wavelength

Suppose the wave in figure 2.1 is moving from left to right at speed c, and we track a given crest. The distance the crest would cover in time t is $d = ct$—distance equals velocity times time. The time between one crest and the next (one wavelength) passing a given point is called the period, τ. In that

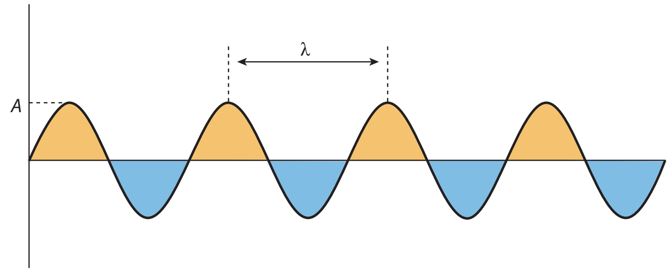


Figure 2.1
Sinusoidal wave with amplitude A and wavelength λ.

case $d = \lambda$ and $t = \tau$, and $d = ct$ becomes $\lambda = c\tau$. The inverse of the period is the number of repeats per second, which we call the *frequency*, $f = 1/\tau$. Then, we can just as well write $f\lambda = c$. The relation

$$f\,\lambda = c \tag{2.1}$$

should never be forgotten. Since $f = 1/\tau$, we can also write

$$\lambda = c\tau. \tag{2.2}$$

This form is simply "distance equals velocity times time" stated in terms of wavelength for distance and period for time.

In air, all wavelengths of sound travel at the same speed. If this were not so, speech would become unintelligible some distance from the speaker, with the higher-frequency, shorter-wavelength components of speech arriving before or after the lower-frequency ones. Not all types of waves obey this equal-speed rule—for example, surface water waves traveling over deep water propagate at a speed proportional to the square root of their wavelength, so at a distance from the source of a sudden disturbance, the long wavelengths arrive first. Anyone who has spent extensive time on the water will have noticed that the longer-wavelength waves from a distant ship arrive before the shorter ones.

It is always a good idea to check the dimensional units of physical expressions. This is called *dimensional analysis*. We want the wavelength to correspond to a length, of course, which means the answer should have dimensions of meters (m). The formula $\lambda = c/f$ is dimensionally correct, since c is a speed, in m/s, and f has dimensions of inverse seconds. (Frequency is the number of events *per second*.) A 1000 Hz tone thus has a wavelength of $\lambda = 343/1000 = 0.343$ m, or about a foot.

The *Ripple* applet solves sound wave propagation numerically, based largely on the "push and pushback" idea. With pixels representing the cells, there is confirmation that the speed of sound propagation does not depend on the wavelength or the frequency of the sound. In figure 2.2, we show the evolution of circular waves emanating from two point sources, one high

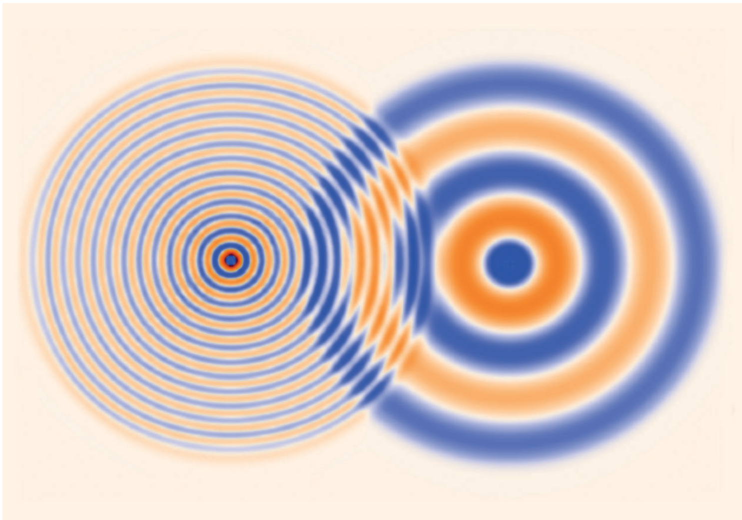


Figure 2.2
The speed of propagation does not depend on the wavelength or the frequency of the sound.

frequency, one low. The two waves started at the same time; the figure shows that they have progressed the same distance.

2.2 Falloff with Distance from the Source

A small tube with an open end in a large space filled with air is a simple source of sound. Small quantities of air can be forced in and out of the tube by connecting it with a piston, for example.

Pressure variations originating at the end of the small tube propagate outward in concentric spheres, which are the three-dimensional analog of the circular surface waves resulting from dropping a pebble in a pond. It is intuitively clear that the pressure variations will become "diluted" and less intense at greater and greater distances from their source. This makes sense because the pressure variations carry the energy required to create them. As they travel away from the source, no new energy is added, so the waves must weaken as they spread. The total energy taken over the expanding sphere remains constant. If $\delta p(t)$ is the *difference* between the air pressure at a point in space and the ambient air pressure at time t, the average $\overline{\delta p(t)^2}$ over some time interval is proportional to the energy of the sound at that point.

Suppose the air flow in and out of the tube is periodic, with frequency $f = 100\,\text{Hz}$, and that it has been flowing for a long time and has spread over a large volume. If we want to conserve energy as it propagates outward, the energy per second passing through the surface of a sphere with the source at its center should not depend on how big the radius of the sphere is. We can visualize this by analogy. Suppose people have been streaming out of a subway station in a park for some time at a constant rate, heading in various directions. We could count the number of people emerging from the subway per unit time by watching how many passed through a small circle surrounding the subway exit, or we could take a larger circle much farther away. Either way, we should get the same answer if people have been coming out at a constant rate for a long time. Of course, the density of people passing through the small circle will be much higher than in the larger one. In fact, the density will drop off as the inverse of the radius of the circle if the people head out uniformly in all directions.

The area of a sphere of radius r is $A = 4\pi r^2$, proportional to the radius squared. Thus, for the total energy to remain fixed, the energy passing through one square meter of surface on the sphere per second must drop off as $1/r^2$. Since the energy is proportional to the square of the amplitude, *the amplitude or pressure* δp *will drop off as the inverse of the distance from the source*. This is a very important conclusion, and it will be found to be true for any source some distance away without obstructions or surfaces nearby.

Loudness Falloff with Distance

Part I of this book discusses objective aspects of sound, measurable by instruments. Nonetheless, we take a short detour here to discuss the issue of the *subjective* falloff of loudness with distance from the source, which is of great practical importance. A more complete discussion of subjective measures of loudness can be found in chapter 22. Sound intensity, I, as measured by instruments, and sound energy are proportional to each other. (We will see how to quantify sound intensity in decibels later.) Both are proportional to the mean square of the fluctuations in pressure: $I \sim \overline{\delta p(t)^2}$. The perception of loudness depends, of course, on sound intensity, but it has also been found to depend on the frequency and duration of the sound. Subjective measures such as loudness, being human impressions, cannot be made truly quantitative, but there is a rule of thumb that seems to work: for a sound to be perceived as twice as loud as before, the intensity of the sound must be increased tenfold. If you want to know how bothersome a noise source will be at some distance away, it is useful to have an approximate measure like loudness N and approximate rules like the tenfold rule just stated.

The tenfold rule specifies that the loudness N and the sound intensity I are related by $N \sim I^{0.301}$, which follows from the fact that $2 = 10^{0.301}$. Solving for I, we have $I \sim N^{3.32}$; for example, $10 = 2^{3.32}$. Note that if the intensity increases tenfold, $I \to 10I$, then $N \to 2N$. Since $N \sim I^{0.301}$, and I drops off as $1/r^2$, we have

$$N \sim I^{0.301} \propto \left(\frac{1}{r^2}\right)^{0.301} = \frac{1}{r^{0.602}}.$$

Again, r is the distance from the source. Thus our impression of sound intensity—that is, its loudness N—is expected to fall off from the source more slowly than the intensity, $N \sim 1/r^{0.6}$, whereas $I \sim 1/r^2$.

If a jackhammer is earsplitting at 10 m, how far away do you have to be before it sounds four times less loud—something you could perhaps tolerate? The reference loudness is $N_0 = A\frac{1}{r^{0.6}} = A/10^{0.6}$, and we want to solve for r: $N/N_0 = 0.25 = 10^{0.6}/r^{0.6}$. Solving for r, we find that we must travel about $r = 100$ m away—a tenth of a kilometer (km)—for the sound to become tolerable. For the jackhammer to stop invading our consciousness, it might have to be 1/16th as loud as it is at 10 meters, which translates to about a kilometer away. Of course, at this distance, it is still quite audible. At distances of a kilometer or even much less, various new factors come into play, including the fact that the sound may be affected by the ground; reflected or shielded or absorbed to some extent by various objects; or refracted by wind, which could make it much louder or much softer. (We will discuss sound refraction by wind and temperature gradients, which can have drastic effects on sound, in chapter 28.) More on loudness can be found in chapter 22.

The decibel is the best way to report sound power, and it is frequently used to describe loudness (see section 2.10).

Ripple Simulation

We can check the falloff in intensity using Paul Falstad's *Ripple* applet, except that it is set up to mimic waves in two dimensions, not three. Going back over the preceding arguments for a circle instead of a sphere, we conclude that the energy will drop off as the inverse of the radius of the circle, and the amplitude will drop off as the inverse of the square root of the radius. We set up *Ripple* with a single monopole source and four detectors, placing them at multiples of a given distance from the source. Using a sinusoidal source (see chapter 3 for more about sinusoids), we monitor the four probes, capturing the data and examining it in the resulting text file that is produced. Figure 2.3 shows the setup and the resulting data from this numerical experiment. We see good agreement with the expected falloff as one over the square root of the distance from the source. This is represented as the smooth curve, and the data points obtained by reading the maximum amplitudes from the *Ripple* data file for the four probes are shown as dots.

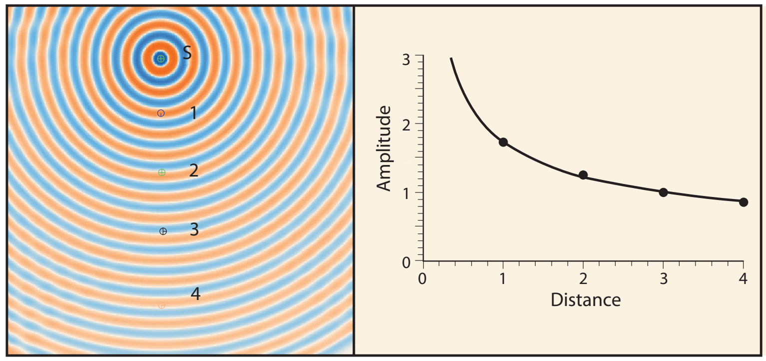


Figure 2.3
Modified Falstad *Ripple* Java applet experiment modeling a point sound source S in two dimensions. Four probes are spaced at approximately equal distances from each other and from the source. This scenario and many others are easy to set up on any computer that can run Java. The color scheme in *Ripple* was changed for this book; blue regions correspond to rarefaction, orange, to compression. The maximum amplitude for each of the probes was recorded and plotted. A curve falling off as one over the square root of distance from the source is also shown. The falloff can be seen on the left as brightness that diminishes as it moves away from the source.

2.3 Measuring the Speed of Sound

As everyone knows from hearing echoes, sound waves *reflect* from objects and surfaces. The reflection is *specular*, or mirror-like (the angle of incidence equals the angle of reflection), if the reflecting object is large and smooth.

The time it takes for an echo to return from a wall at a known distance could be used to determine the speed of sound. Why didn't the best scientific minds in the last two thousand years determine the speed of sound using echoes? One problem was the accurate measurement of short intervals of time. But for centuries, it wasn't clear what actually happened when sound reflected from a surface. Did the surface vibrate in order to send the sound back? Was there a time delay with the sound held at the wall for an interval of time?[1]

Confusion about the nature of sound reflection from hard surfaces thus seems to have postponed what should have been an obvious way to measure the speed of sound. The French ecclesiastic Father Marin Mersenne (1588–1648), of the order of Minims, made the first good estimates of the speed of sound using echoes. The method for measuring short intervals of time favored by Mersenne was speaking polysyllabic words in rapid succession: Mersenne could utter seven syllables per second. This timing method, together with the echo from a distant object, led to a figure of 1038 ft/s for the velocity of sound, not too far off from the 1125 ft/s modern value at room temperature, and very close the correct value at 32°F, or 0°C.

[1] The existence of a time delay could have been checked in several ways, including clapping one's hands close to a wall.

Box 2.1

Father Marin Mersenne

Father Marin Mersenne (figure 2.4) is sometimes called the father of modern acoustics. His experimental methods were remarkable for their day, as documented in his 1636 treatise, *Harmonie Universelle*. In this work, he spelled out the Mersenne laws (see chapter 8), the connections between length, tension, diameter, and frequency of a string. Mersenne established that an octave was an exact 2:1 ratio of vibrations per second. Starting from oscillations so slow he could count them accurately, Mersenne ascended by exact octaves, and in this way was the first to associate a specific number of oscillations per second (84 exactly) of a string with a given audible pitch. Galileo (1564–1642) may have actually beaten Mersenne to many of his discoveries, but Mersenne was first in print.

Mersenne possessed a remarkable gift for networking with other scientists (and their work) from far-flung places. He became a kind of one-man clearinghouse for scientific information. Many of the most famous scientists of the time corresponded and visited with Mersenne. French philosopher René Descartes (1596–1650) and mathematician Blaise Pascal (1623–1662) met each other for the first time in Mersenne's cell at the monastery. Upon Mersenne's death in 1648 from a lung abscess, letters from 78 correspondents, including Fermat, Huygens, Pell, Galileo, and Torricelli, were found in his cell, along with various scientific instruments and unfinished manuscripts.

Mersenne was the founder of the French Academy of Sciences, which began with his prodigious correspondence and frequent meetings at his Paris cell. He was also a staunch defender of Galileo in the latter's travails with the Catholic Church, even though Galileo was his strongest scientific competitor. The list of Mersenne's talents continues: he was a fine musician and a world-class mathematician. The Mersenne prime numbers, as they are called, are a lasting reminder of his mathematical discoveries and talents.

Figure 2.4
Minim Friar Marin Mersenne (1588–1648).

2.4 Interference and Superposition

The circular pattern of compressions and rarefactions spreading out from a point source at the speed of sound is shown in figure 2.3. If the source had been in a different position, the image would have been exactly the same, except the concentric wave disturbances would have the new position as their center. Imagine we take the trouble to record the wave amplitude everywhere at a given time for each of the two positions run separately. Next we run with both sources active at the same time, and again we record the amplitude everywhere. We find that the sum of the two amplitudes from the individual sources agrees perfectly with the amplitude found by

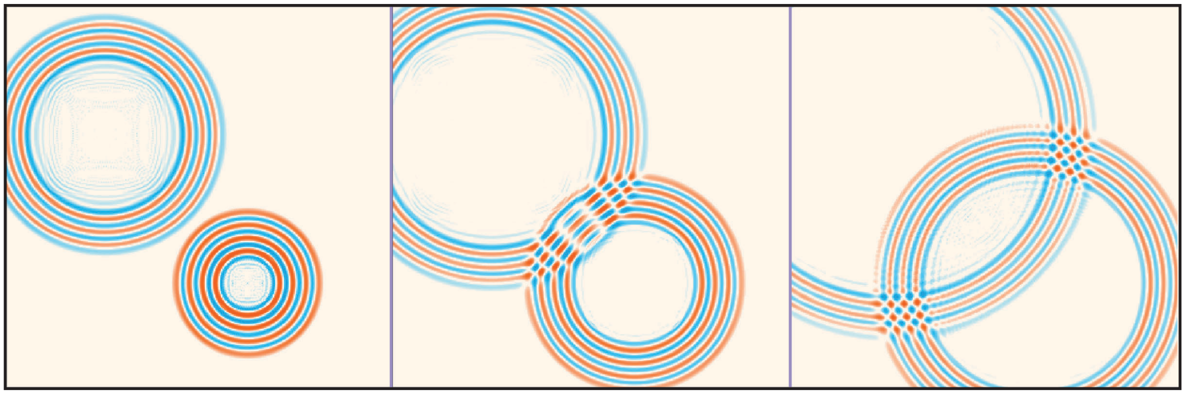


Figure 2.5

Two pulsed sound waves emanating from point sources: before first contact, just after first contact, and sometime later. Notice that even though the waves interfere in overlapping regions, they pass right through each other.

running both of the sources simultaneously. This is an example of the principle of superposition: Suppose you have a wave defined by a pressure field $\psi_1(r, t)$, which might be a sound wave in a room at position r and time t coming from one of two stereo speakers, with the other speaker disconnected. Then consider a totally different wave $\psi_2(r, t)$ coming from the second speaker playing a different tune from a different place in the room with the first speaker disconnected. We may then merely add the two waves, as $\psi(r, t) = \psi_1(r, t) + \psi_2(r, t)$, to determine what would happen at every point in the room if both speakers are playing their respective tunes at the same time.

The superposition principle is a consequence of the fact that sounds pass right through each other, one not disturbing the other. A little creative experimentation in Falstad's *Ripple* applet will convince you of this. Life would be very different if this were not true. The sound at a loud party is the sum of the sounds of all the individual sound sources obtained *as if they were acting alone*. If someone is speaking to you from across the room, you still hear other conversations in the room, of course, but they are no louder or softer than if that person had not been speaking to you. Light waves, matter waves as in quantum mechanics, and, to a good approximation, water waves are the same, obeying the superposition principle.

Even though sound waves pass through each other unscathed, they locally interfere with each other destructively or constructively, making the sound softer or louder. We must take the equation $\psi(r, t) = \psi_1(r, t) + \psi_2(r, t)$ literally; to find the pressure at r, we have to add the pressure from ψ_1 and ψ_2. If at that moment t and at that place r we have $\psi_1 = 0.001$ and $\psi_2 = -0.001$, then the two cancel (destructively interfere), and $\psi = 0$. These points are made clearer in figure 2.5, which shows two localized sound waves before, during, and after they occupy the same space. Interference takes place in the region of collision between the two expanding rings, but the collision has no effect on either ring as they propagate beyond the point of intersection. At every moment and at every place, the total disturbance is simply the sum of the individual disturbances.

Active Noise Cancellation—Deliberate Destructive Interference

Active noise cancellation technology is now common in audio headsets. The idea behind this technology is to use external microphones to capture the sound on its way to the ear and, using fast electronics, have the headset speakers apply an opposing pressure field just in time. The principle of superposition does the rest, in theory, canceling the noise. In practice, the cancellation is not perfect, and the sound power can be reduced by about a factor of 100, or 20 dB, in certain ranges of the audio spectrum. Achieving a better result than this is difficult for several reasons, including direct bone conduction of sound to the middle ear. In noisy environments, plenty of noise enters your ears in spite of the noise cancellation. Contrary to rumor, noise cancellation earphones are not degraded for safety reasons, although that might be a good idea if they could in fact attenuate by 40 or 60 dB.

Use of noise cancellation technology can be a huge boon to your hearing, or rather the survival of your hearing into middle and old age: with the noise reduction, you don't have to turn the music up as high to overcome ambient noise so the total sound level remains reasonable. The sound file Jet Cabin Noise Seat 12A on whyyouhearwhatyouhear.com demonstrates cabin noise in a Boeing 757 at 30,000 feet, and the same noise reduced in volume by 18 dB, about the amount for average quality noise-cancellation headsets, for comparison.

High frequencies are problematic for noise cancellation headphones. There are at least two reasons for this. First, the electronic circuitry and the headphone speaker have to react very fast to changes in sound pressure at high frequency. Second, high frequencies (say, above 3000 to 4000 Hz) start to have wavelengths comparable to the interior of the headphones, making cancellation more difficult, since the sound pressure becomes different in different places inside the headphones. In principle, the best cancellation technology is achievable with earbuds that seal off the outer ear: one can combine passive cancellation (itself good for about a −20 dB attenuation) and active cancellation.

2.5 Reflection

When you look in a mirror, you see light that has been *specularly* reflected, bouncing off the mirror according to the following rule: the angle of incidence equals the angle of reflection (figure 2.6). If the surface gently

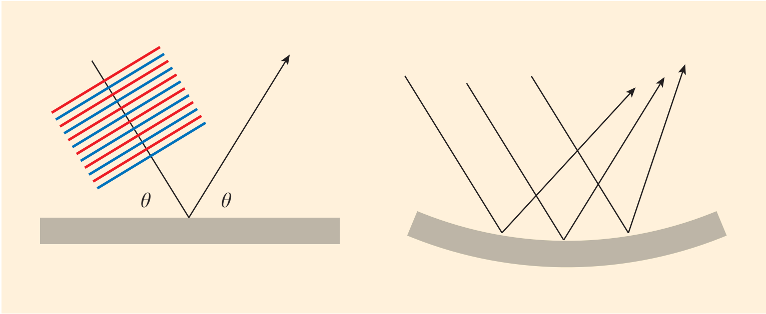


Figure 2.6
Wavefronts (crests shown in orange and troughs in blue) approach a flat surface. The ray path corresponding to the propagation of the energy in the wave is shown. The ray is perpendicular to the wavefronts and bounces specularly off the wall. On the right, parallel rays fall on a smooth but curved surface; the locally specular bounce leads to rays reflecting at different angles.

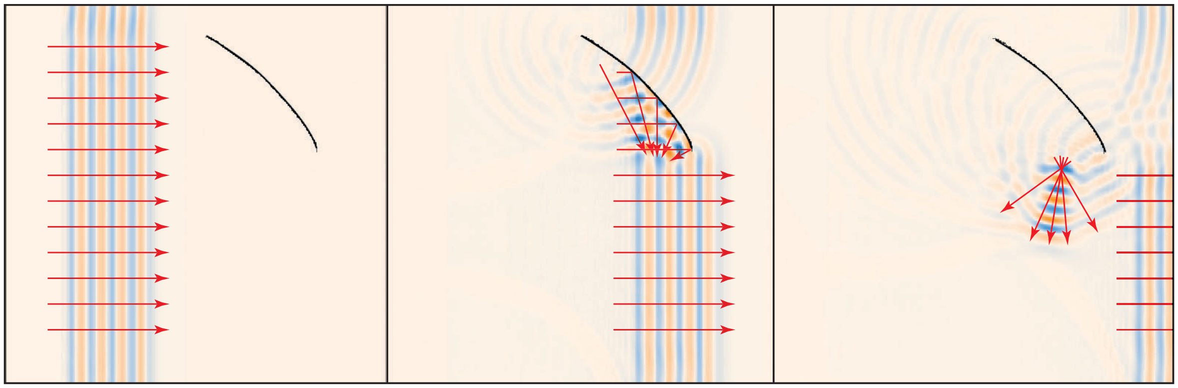
curves on a scale much bigger than the wavelength, the angle of incidence rule holds locally at each point along the surface. Initially, parallel rays fall at different places on the surface and are directed at various new angles relative to each other, as in a concave telescope mirror.

Figure 2.7 depicts a simple simulation of reflection and focusing due to a curved sail.

Shiny and Matte

When light falls on a matte-painted wall, it reflects *diffusively*, scattering it in many directions, leaving no sharp images of reflected objects. Shiny surfaces scatter light *specularly*. However, *every* real surface is quite rough

Figure 2.7
Reflection and concentration of waves due to a concave surface in a *Ripple* simulation of the sound-concentrating potential of a curved sail, as discussed at the beginning of this chapter. Here, a short burst of sound is shown progressing from left to right in the three frames; a concentrated (focused) beam of reflected sound is seen in the frame on the right. The ray tracing of the situation is depicted as red arrows; the convergence of the arrows reveals the focal point, seen also as a concentration of the wave amplitude.



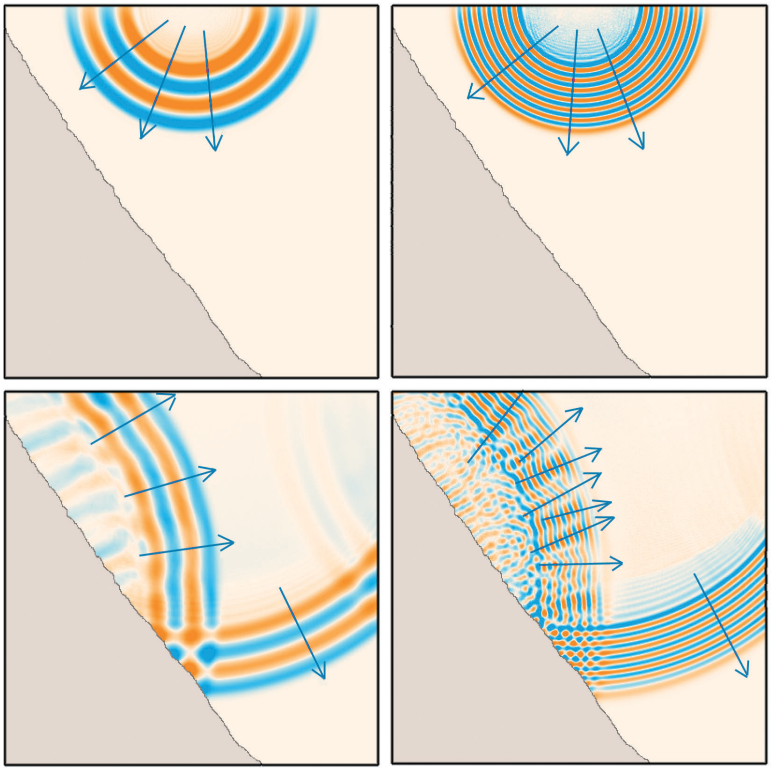


Figure 2.8
Long- and short-wavelength sounds impinge on a surface with small-scale roughness; much more “damage” is done to the short-wavelength sound, even though the surface is the same in both cases. The short wavelengths reflect in a nonspecular way, although in the situation shown here they have not completely forgotten their initial direction, and the surface might appear “semigloss” at this wavelength.

if one looks with enough magnification, so why do surfaces ever look shiny? The key issue is whether the surface roughness is on a scale large or small compared to the wavelength of the impinging wave. If the roughness is small, on the scale of a wavelength, the wave bounces off as if no roughness existed. The wave can’t “resolve” features smaller than its own wavelength, and averages over them (see figure 2.8).

If the surface is instead rough, on the scale of or somewhat larger than the wavelength, the reflection becomes diffuse, meaning the waves head off in a range of directions, not just in the specular direction. For example, a brushed metal surface reflects light rather diffusely because the wavelength of visible light is smaller than the imperfections and grooves in the metal. The same surface will be mirror-like for microwaves with wavelengths on the order of centimeters. Similarly, an ordinary bathroom mirror is not in fact perfectly smooth, but rather just smooth enough—the surface of the silvered or aluminized glass that forms the mirror is quite rough, but the scale of the roughness is small compared to a wavelength of visible light.

These statements apply to sound waves reflecting from a hard surface. A surface with only small irregularities but otherwise flat reflects low-frequency, long-wavelength sound in a mirror-like fashion. The same surface will reflect high-frequency, short-wavelength sound diffusely. These facts have not been lost on designers of acoustic spaces, especially concert halls. The effect of the scale of the roughness compared to the wavelength is illustrated in the *Ripple* simulation shown in figure 2.8. A point source some distance from the surface sends long- and short-wavelength sound toward a surface; much more "damage" is done by the rough surface to the short-wavelength sound, which reveals clumps of waves traveling in nonspecular directions.

2.6 Refraction

Wave energy often progresses in a well-defined direction, but that direction can change more or less slowly. Such *refraction* often goes unnoticed for sound waves, but refraction is actually quite ubiquitous outdoors over distances of about 100 m and beyond. Refraction results from the variation of the speed of the wave within the medium in which it is traveling; the bending or curving of the wave is always toward regions of slower wave speed.

Several factors cause the wave speed to vary in air, all understandable in terms of the concepts introduced in chapter 1: the drunken messenger model, the air cell impedance picture, or both.

Temperature. Lower temperature means the molecular messengers are moving more sluggishly, reducing the speed with which pressure fluctuations are propagated. Every gas atom or molecule has on average the same energy as its neighbors, independent of its mass. Energy is defined as $E = 1/2\, mv^2$, where m is the mass of the molecule, and v is its velocity. Energy per molecule in a gas is proportional to temperature, expressed in kelvin, or K (room temperature is 295 K). Thus at 273 K, the freezing point of water, the speed of sound in any gas should be $\sqrt{273/295} = 0.965$ times as fast as it is at room temperature, 295 K. At 295 K, the speed of sound in air is 343 m/s; thus we predict it to be $343 \times 0.965 = 331$ m/s at 273 or 0°C. This is indeed the measured value.

Composition. Differences in chemical composition change the messengers themselves. At the same temperature, lighter messengers are speedier. Again, the energy of any molecule due to its speed is the same as that of any other molecule (this is called *equipartition of energy*), and since $E = 1/2mv^2$, the average speed must be higher if m is smaller. The speed of sound in air, a mixture of nitrogen (about 80%) and oxygen (about 20%) with an average mass of 29 grams/mol is 343 m/s at 20°C (room temperature). As discussed earlier, we would expect the sound speed in helium gas, mass 4, to be about $\sqrt{29/4} = 2.7$ times faster than air, or $343 \times 2.7 = 929$ m/s. The measured value is 972 m/s. Sulfur hexaflouride, SF_6, should have a speed of $343 \times \sqrt{29./146} = 153$ m/s; the measured value is 150. In the atmosphere, water vapor content is the most common cause of composition changes from one place to another.

Motion of medium. Last, if in some region the messengers are moving en masse in the same direction, the wave propagates slower or faster (over the ground) according to whether it is moving with or against the mass movement. This will speed up or slow down the wave arrival merely by a fraction, except if the speed variation differs from place to place, in which case the variations *also cause refraction of the waves*. Temperature, composition, and speed gradients are common factors affecting sound outdoors; they will come up again in chapter 28.

A way to quickly (if qualitatively) follow wavefronts to see how sound (or light) propagates was invented by Christian Huygens more than 300 years ago. This is a third way of understanding sound propagation, in addition to the drunken messengers model and the cellular method (chapter 1).

Huygens's method works as follows: we start with a wavefront representing some wave incident on a "scene" that may include different materials. We want to construct the wave farther along the direction of propagation. Along the initial wavefront, we locate the centers of arcs of constant radius; the "envelope" of the new arcs is the new wavefront, also of constant phase. If the arcs are half a wavelength in radius, the new wavefront will be a crest if the old one was a trough. To understand refraction, we slightly modify Huygens's original illustration. In figure 2.9, a wave is traveling in the direction of the line segment D-A, with perpendicular wavefronts—for example, A-C and the lines K-L. Suppose τ is the time it takes for the wavefront to travel the distance from C to the leftmost L. This distance is the radius of the Huygens arcs used to propagate wavefronts in the upper medium. (We have drawn a few of these in red.) Along the interface between the two media, the wavefronts will advance between the adjacent points labeled K in a given time τ. Inside the medium indicated by the rectangle, the speed is lower, and we draw correspondingly smaller radius arcs. The arc whose center is the rightmost point labeled K has such a smaller radius and represents the progress of part of the wavefront from that point in one unit of time τ. Huygens has drawn an arc of twice this radius from the adjacent point K to its left, representing a time 2τ since the wavefront entered there, and three times the radius from the point to the left of that. This is the correct procedure, as can be seen by the intermediate wavefronts inside the medium (colored blue) given by the lines labeled K-O. The new wavefronts are not parallel to the old wavefronts outside and above the medium. We have thus constructed the new wavefront inside the medium. Employing the rule that the energy flow is perpendicular to the wavefronts, we see that there is a new direction A-N inside the medium compared to the old direction D-A of propagation above the medium. We have shown that the wave refracts as it enters the medium of slower wave speed. Notice that the ray bends *toward* the medium with slower wave speed—this is a useful rule to remember about refraction.

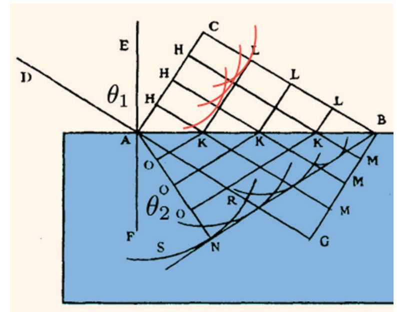


Figure 2.9

Christian Huygens's depiction of the geometry of refraction (from Huygens's *Traité de la Lumiere*, 1678), as determined by his construction using arcs to advance a wavefront, with the addition of the red arcs.

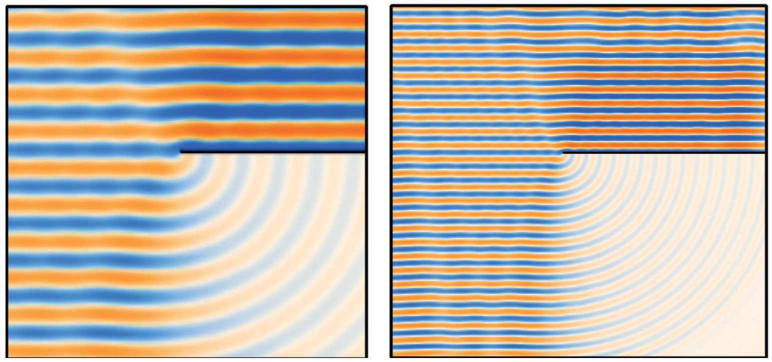


Figure 2.10
A plane wave having long wavelength arrives from the top of the image and is interrupted by a reflecting wall (left). The geometric shadow is a vertical line heading straight down from the end of the wall. Waves penetrating beyond this line into the shadow region are by definition "diffracting." More diffraction occurs for longer-wavelength (lower-frequency) sound; the diffracted power is proportional to the wavelength. This figure is taken from a *Ripple* simulation; claims about the amount of diffracted intensity can be checked by setting up probes at appropriate positions.

If you are sitting inside the lower medium, the wave crests arrive one after the other with the same period as in the medium above. (If a person above the surface of a pool is waving her arm once per second, the period will be one second whether you are looking from above or below the surface.) Because the wave is moving more slowly in the lower (blue) medium, the wavelength must be shorter to keep the frequency f the same. The rule is $f = c_1/\lambda_1 = c_2/\lambda_2$, where the c's are the wave speeds and the λ's are the wavelengths in the two media. It is then quite simple to show geometrically, using the fact that the wavefronts must agree at the interface, that

$$\lambda_2 \sin\theta_1 = \lambda_1 \sin\theta_2, \quad (2.3)$$

where the angles θ_1 and θ_2 are the angles of incidence and refraction indicated in figure 2.9. This is Snell's law, derived in many elementary physics textbooks, usually in the context of light waves. There are many practical implications for the refraction of sound that we will encounter along the way in this book.

2.7 Diffraction

Diffraction is another way that waves, including sound, can take a non-straight-line path. For projectiles launched from a distant point, a solid object forms a "hard shadow," a refuge from the projectiles. Waves do not respect such hard shadows; they can bend around edges of obstructions, making it possible to "hear around corners." Diffraction is less pronounced for short wavelengths, so if a marching band is approaching out of sight behind a building, the low thumping of drums will be heard well before the piccolos. This fact is reinforced by the simulations in figures 2.10 and 2.11.

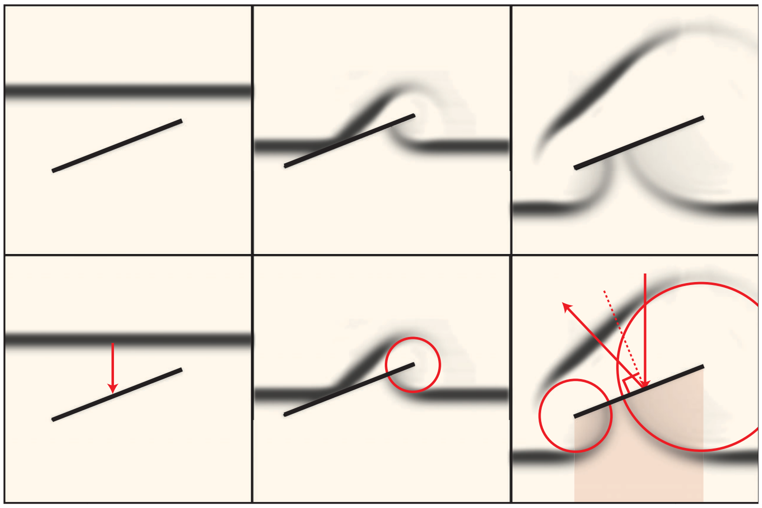


Figure 2.11
(Top) Successive snapshots of a sound pulse arriving from above a wall segment. (Bottom) Diffraction occurs from the end points and is similar to a pulsed wave from those points leaving as the initial pulse arrives, as indicated by the red circles. The analysis shows that the main portion of the reflected wave obeys the rule angle of incidence = angle of reflection. The "hard shadow" region, shaded pink, would be free of sound if sound traveled in straight lines, without diffraction. Taken from screenshots in a simulation run in *Ripple*.

Figure 2.10 shows diffraction around a wall segment for two different wavelengths of sound. We see that long wavelengths diffract more easily, or put another way, there is a closer approximation to a hard shadow for very short wavelengths.

Diffraction at an Edge

We may visualize diffraction due to an edge in terms of the cell picture of sound propagation. If we examine figure 2.10 closely, it is apparent that the end of the wall is very much like a point source of sound. A circular wave emanates from that point, as is especially evident in the lower-right quadrant, where nothing else is present. Faint evidence of this circular wave can also be seen in the other three quadrants. See also figure 2.11.

The diffraction from the edge is understandable from the cell model as follows: a cell just above the wall, but more than about half a wavelength from the end point of the wall, won't "feel" the end point, since it is too far away to send and receive information (within one period) from the end point at the speed of sound. The impedance of such a cell is therefore not reduced because it does not sense the lower resistance to pressure changes near the wall. More than half a wavelength beyond the end of the wall also acts like any other cell in free space, again because it is too far away to send and receive the information that the wall is present. There is one special cell acting strangely, straddling the end of the wall. This cell is busy scattering sound because its impedance is different than any adjacent cell. It scatters whatever impinges on it. The scattered (diffracted) wave may be understood as coming from this small region, which therefore acts like a small, or *point* source of sound. Like any source, there will be a falloff in amplitude as distance increases from it; in this two-dimensional example, we have the diffracted amplitude $a_d(r)$ falling off as

$$a_d(r) \propto \sqrt{\lambda/r}, \tag{2.4}$$

and the intensity thus declining as $\propto \lambda/r$. The proportionality involves only factors of 2 and π. The power passing through the cell of length λ is proportional λ—it's a bigger cell at longer wavelength. The amplitude squared of the diffracted wave is proportional to the power, so the factor of $\sqrt{\lambda}$ in the numerator correctly accounts for the *amplitude* passing through one cell. As the wavelength gets larger, the odd cell at the end, which is one wavelength across, diffracts more power in proportion to its wavelength.

Brush with the Law of Similarity

For a thin wall, the scale of the picture is set by the wavelength. The left panel in figure 2.10 is about 10 wavelengths across in both directions. If someone declares that the frame is physically 100 m across, then the wavelength is about 10 m; from $f\lambda = c$, the frequency is about 34 Hz. If the frame is only 10 m across, the wavelength is about 1 m and the frequency is 344 Hz. *The picture is correct either way.* By this principle, a mockup only 1 meter high could be used to study the diffraction of sound around a highway barrier that in reality is going to be 10 m high, *provided all the sound frequencies are increased correspondingly by a factor of 10*. This is our first encounter with the law of similarity, which allows us to scale up and scale down studies of wave propagation, diffraction, and so on by scaling the physical dimensions such as wavelength λ by some factor such as 1/10, and scaling the frequency by a factor of 10 *so that* f λ = c *both before and after the scaling*.

Similarity implies that sharp edge diffraction is always the same: you need only one picture! It's just a matter of scaling the picture up or down. But then what is the justification for claiming that long wavelengths diffract more than short ones?

We have already shown why the diffracted power increases in proportion to the wavelength—that is, more is diffracted for lower frequencies—the *drum and piccolo effect*. The similarity argument we are now making reinforces this: measuring distances in a picture such as figure 2.10 in terms of wavelengths, not meters, the amplitude $a_d(r)$ in equation 2.4 falls off at the same rate. For example, suppose r is 10 wavelengths away—that is, $r = 10\lambda$. If the wavelength is 10 m, r is 100 m from the wall; if the wavelength is 1 m, r is only 10 m from the wall. This confirms that there is more diffraction in the case of the longer wavelength, *since it has the same sound intensity in the shadow region 100 m from the edge as does the shorter wavelength 10 times closer to the edge, at 10 m.* This also confirms with the notion that a region about one wavelength wide is responsible for the diffraction. The longer the wavelength, the more energy is thrown into diffraction, since power passing through an opening 10 m wide is 10 times that passing through an opening 1 m wide. The law of similarity is taken up in more detail in section 7.6.

Active Noise Reduction of Diffracted Sound

Long highway sound barriers are now routinely placed between traffic and residential areas, although they are quite expensive—about two million dollars per mile. The walls are usually very solid, so most of the sound arriving at the houses (if the wall blocks the line of sight with the traffic) must have been diffracted. We now know this sound comes from the top edge—that is, it is diffracted near the top of the wall. The amplitude falloff will be a function of distance from the top of the fence.

This makes possible an "active" sound attenuation strategy, namely, to *put an out-of-phase source right where the culprit "edge diffraction source" lies*. This can be accomplished by mounting microphones and loudspeakers along the top of the sound barrier. Sound impinging on the edge is detected by the microphone, processed by a computer chip and re-emitted *out of phase with the incident sound* by loudspeakers aimed toward the quiet side of the barrier. The speakers could be efficient horn loudspeakers (see section 7.3) and solar powered. The speakers and microphones need to be placed densely along the wall, but compared to two million dollars per mile, it might prove cheap if the attenuation worked well enough.

A *Ripple* simulation of a similar situation is shown in figure 2.12. A space between two vertical walls, with no roof, is filled with loud sound due to

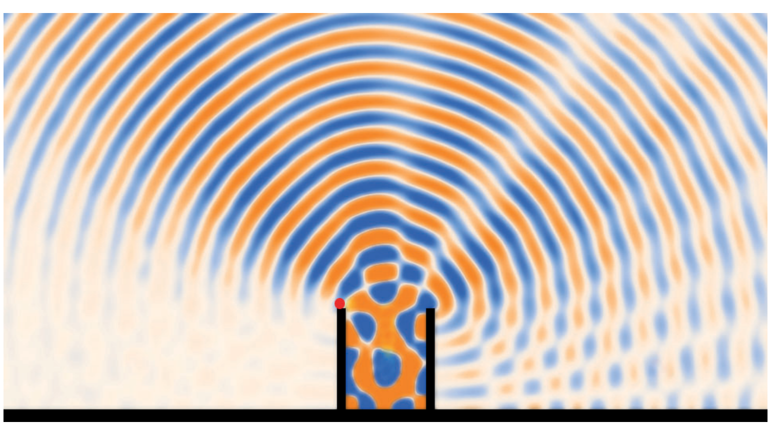


Figure 2.12
Ripple simulation of active noise cancellation from a noisy area (between the walls), using destructive interference. The red dot is the location of a point source out of phase with the diffracted wave, which is for the most part successfully canceled on the left.

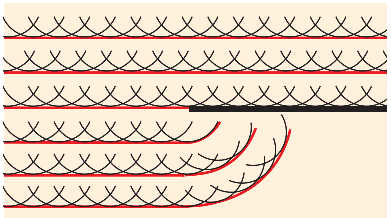


Figure 2.13
Huygens's wavefront construction of the diffraction that results when a plane wave collides with a wall. The wavefront constructed from the wavelets (black curves) becomes curved near the end of the wall. Subsequent applications of Huygens's rule leads to the propagation shown and the development of the curved diffraction wavefront. Using Huygens's wavefront construction, it is difficult to quantify the amplitude in the diffractive region.

seven sources. An eighth source is placed at the top of one wall; its phase and amplitude are such as to destructively interfere with and therefore attenuate the diffraction reaching the ground, as seen on the left; compare this with the diffraction reaching the ground on the right, for which no cancellation was used. This scenario takes full advantage of the fact that the diffraction from an abrupt edge is itself like a point source.

Diffraction may be understood qualitatively within the Huygens construction. Suppose a wave is incident on a segment of a wall. This leaves a shadow region that however is partly filled with diffracted waves—that is, waves that have deviated from the linear path they were on before they hit the wall (figures 2.11 and 2.13).

To develop an intuition for reflection and diffraction of sound from various objects of different sizes and shapes, it is recommended that you set up various *Ripple* scenarios, drawing obstacles, baffles, objects, and so on and observe the reflection and diffraction of waves of various wavelengths sent at them.

2.8 Schlieren Photography

Schlieren is the German word for optical inhomogeneities in transparent material. Schlieren photography, which makes even slight inhomogeneities visible, was invented by the German physicist August Toepler in 1864. He succeeded in photographing shock waves in air created by *supersonic* (faster than the speed of sound) objects. A sharp *pressure pulse* (shock wave) wave can be created by an electrical spark; we hear this wave as a sudden "snap."

If light propagates through regions of rapidly changing air density, there is a slight deflection toward the denser regions, normally too small to be noticed. In the correct circumstances, especially when the light has to travel large distances to the camera or the eye, small deflections can build up a large effect. Almost everyone has seen "heat waves," the wiggly distortion of objects caused when light passes through heated and disturbed air. Warmer air is less dense and has a lower refractive index than cool air. Cinematographers have a favorite trick for showing something a long way off on a hot day, capturing the wavy distortions caused by refractive index differences of pockets of warmer and cooler air. Light traveling through the turbulent, variable-index medium has a characteristic scintillating and mottled appearance, owing to the schlieren effect. The effect is quite noticeable near very hot objects—for example, near a stove or a candle, where the air density is dramatically reduced due to heating.

Toepler had the idea that a point source of light might be refracted enough by such disturbances to cast lighter and darker bands on a screen

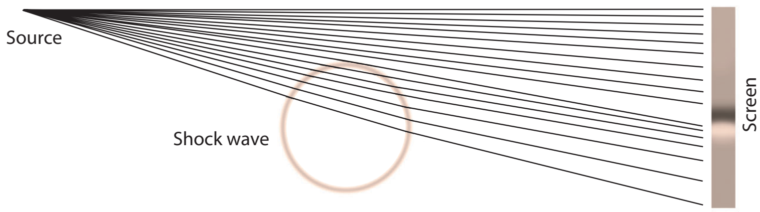


Figure 2.14
Principle of schlieren photography. Slight variations in the index of refraction of air, caused by a propagating circular shock wave (brown), refract light rays. The speed of light is slightly slower in denser air, causing light to bend toward denser regions. Those rays that graze the higher-density disturbance are refracted toward them, deflecting them slightly. This slight deflection is enough to have an effect on the rays cast on a distant screen, causing light and dark bands: light where the extra rays arrived at the screen; dark where they are missing. The rays that graze the shock wave tangential to it are deflected most because they spend the most time near the gradients in the density of air.

some distance away. This works very well indeed, and when the disturbance is localized it can give a very accurate image of it. The principle is illustrated in figure 2.14. Figure 2.15 demonstrates schlieren imaging of a shock wave traveling through tubes of different shapes.

2.9 Ray Tracing

Ray tracing has been around a long time as a substitute for having to solve for the exact wave motion. For example, in describing his speaking trumpet, Sir Samuel Morland used as evidence for the efficacy of ray tracing a pewter parabolic mirror that had not only set a board on fire upon focusing the sun's rays on it, but had also focused a distant man's voice to the same spot (presumably as the shadow of the speaker's head fell across the mirror). Light or sound, ray tracing the waves comes to the same conclusion: they focus at the same spot. A wonderful bit of science for its time.

Optical devices are designed by ray tracing because following the wave motion is far too expensive. A simple rule is used: Rays travel in straight lines unless interrupted by walls or lenses, following the course of the

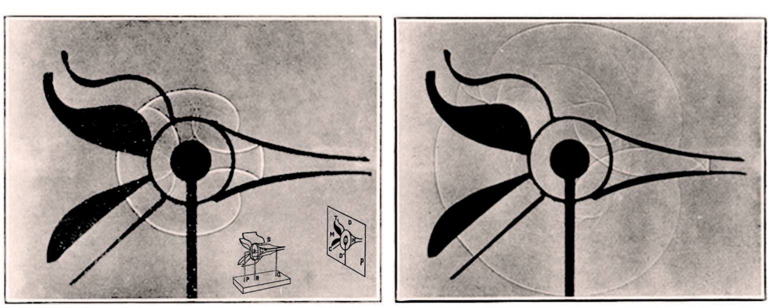


Figure 2.15
In an image created by Arthur Foley and published in *Physical Review* in 1922, sonic (traveling at the speed of sound) shock waves caused by a spark are reflected and guided by three different-shaped tubes. One tube is straight, one is curved, and one decreases in diameter away from the spark. The inner disk hides the spark; the ring surrounding the disk supports the three tube-shaped enclosures and does not lie in the same plane as the spark and does not disrupt the shock waves. The black part of the image is a projection of a three-dimensional object onto a plane (see inset). From Arthur Foley, "A Photographic Study of Sound Pulses between Curved Walls and Sound Amplification by Horns," *Physical Review* 20 (1922), 505–512. © 1922 The American Physical Society.

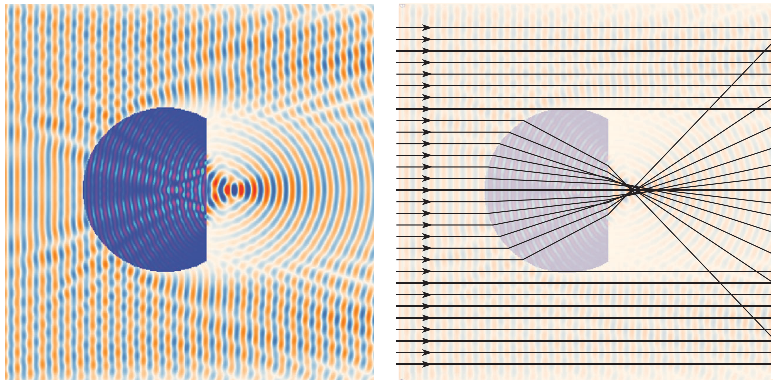


Figure 2.16
Simulation in *Ripple* of a plane wave incident from the left onto a truncated circular lens. On the right, a ray tracing analysis is shown, using only the rays that have not reflected from interfaces.

energy emitted from the source. More precisely, the rays travel perpendicular to the wavefronts, for which they are the surrogates.

At interfaces like that between air and glass, the impedance mismatch tells us that part of the wave is reflected and part is transmitted. This is not a problem for ray tracing: the wavefronts (and with them the rays) split into a transmitted part and a reflected part at interfaces such as air and glass. The part of the wave that penetrates the glass changes direction unless it is incident perpendicular to the interface (refraction). The reflections obey the rule that angle of incidence equals angle of reflection, and the refraction of rays follows Snell's law (equation 2.3). The ray tracing can get quite complicated after several successive encounters with curved surfaces, but it is still much simpler than following the waves themselves.

Ray tracing is only an approximation. It misses diffraction altogether and is accurate and practicable only when the changes in the impedance (and thus the wave speed) in the medium are either very abrupt, in which case there is ray splitting involving reflection and refraction at the interface, or quite slow on the scale of a wavelength, in which case the wavefronts and associated rays curve gracefully.

Figure 2.16 displays a simulation in *Ripple* of a plane wave incident from the left onto a truncated circular lens. On the right, a ray tracing is analysis is shown, using only the rays that do not reflect from interfaces. The full wave simulation shows the effects of interference of the various reflected portions of the waves and the direct waves, as well as diffraction effects.

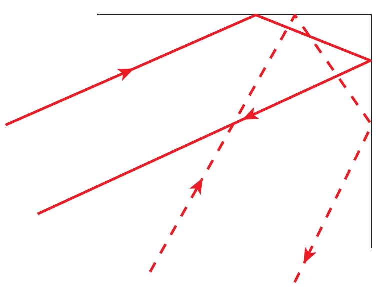


Figure 2.17
A right-angle corner showing two ray paths that bounce once from each wall, returning in a direction exactly opposite to their incident direction. Any ray that bounces from both walls will return exactly parallel to its source.

Corner (Retro-) Reflector

A 90-degree interior corner—three perpendicular walls that meet at right angles—has the interesting and useful property that sound incident on it gets reflected back along the direction it came from, over a wide range of incident angles. This is exactly true in ray tracing analysis, as shown in figure 2.17 for the two-dimensional case. For the three-dimensional

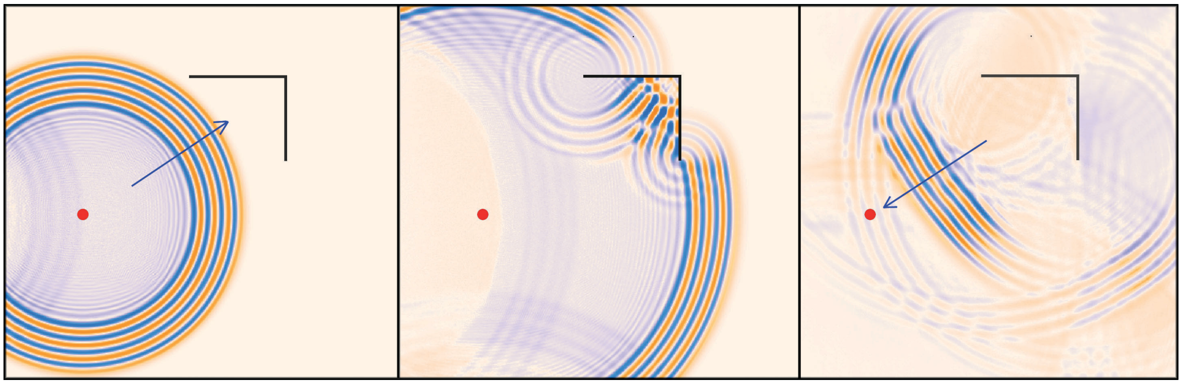


Figure 2.18
Part of a sound pulse from an omnidirectional source (red dot) encounters a corner reflector, which sends some of the pulse back toward its source. In addition, the two weak circular waves seen in the middle frame are the result of diffraction from the tips of the walls (see the discussion for figure 2.19), while the two rather strong side pulses are generated when part of the wave bounces off one wall but misses the opposite wall. The part that hits both walls is reflected back parallel to the direction in which it arrived.

case—that is, the interior corner of a cube—rays can bounce three times, depending on their initial direction, and in every case they return parallel to their incoming path. The sender of a laser pulse will receive a pulse in return. One of the principles of stealth aircraft design is to absolutely avoid right-angle metallic corners that could make the aircraft light up enemy radar screens. On the other hand, small corner radar reflectors are used for boats to make them easily visible on other ship radars. For wavelengths smaller than the dimensions of the reflector, reliable echoes (retroreflections) will be obtained from a corner reflector. A simulation using wave propagation is shown in figure 2.18. The principle has seen wide use, primarily in optics. There are working retroreflectors on the moon, placed there in 1969 by the *Apollo 11* astronauts. Illuminated with laser pulses from the earth, the reflected signal is so strong that the distance to the surface of the moon can be determined very precisely by timing the return pulses.

Sometimes the exterior of buildings will have balconies or other structures that (accidentally) form excellent retroreflectors. The late Professor Frank Crawford of the Physics Department at Berkeley noticed this on the exterior balconies of Latimer Hall on the Berkeley campus. The balconies returned an excellent echo to the sender from a range of different directions.

We conclude our discussion of waves propagating and interacting with different objects with two complex yet informative examples. In figure 2.19, we use a scenario run in *Ripple* to illustrate reflection, refraction, diffraction, and interference, all plainly visible after the wavefronts collide with a block of material with a slower sound speed (which could represent a colder mass of air, or perhaps a heavier gas such as sulfur hexaflouride).

Figure 2.20 shows three different versions (ray tracing, numerical simulation, and hand drawing of an experiment involving liquid mercury) of the result of an off-center source of waves confined to a circular pool.

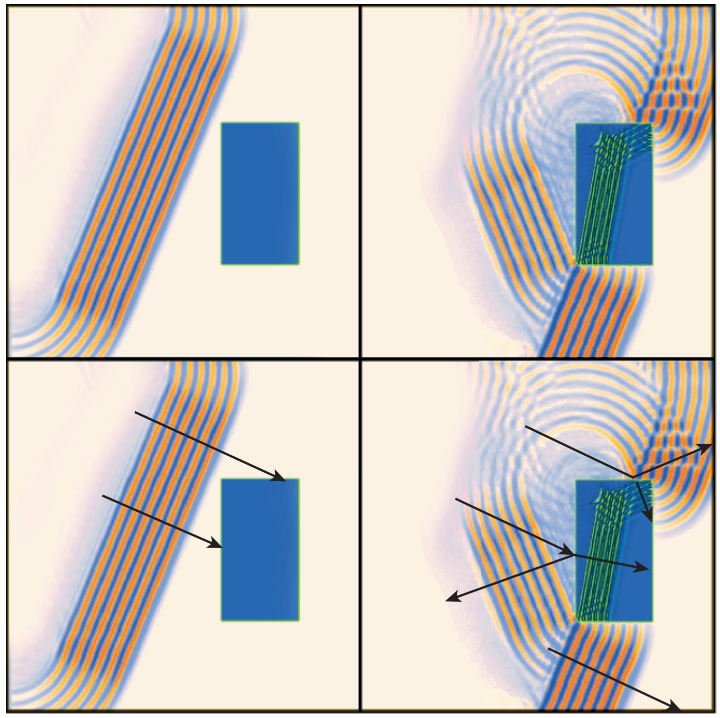
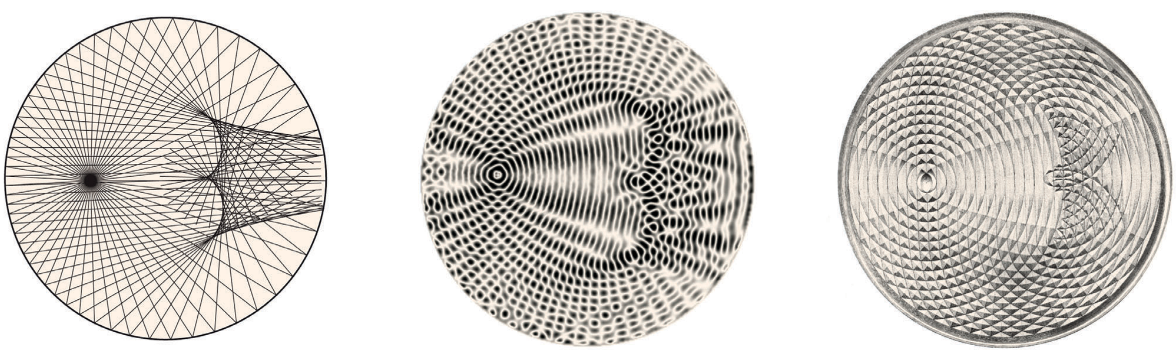


Figure 2.19

A plane wave pulse incident from the upper left travels toward the lower right, in accord with the rule that the energy progresses in a direction perpendicular to the wavefronts. Some ray paths are traced out at the bottom. Reflection, refraction, diffraction, and interference are all plainly visible after the wavefronts collide with a block of material with a slower sound speed. Note the delay in the progress of the wave through the block of material. Also, the wavelength is shorter inside the block. The period must be the same inside the block as it is outside. (Whether you are inside or outside the block, a pulse sent at 1 Hz would have to be received at 1 Hz.) Since the period is the same but the speed is slower, $\lambda = c/f$ implies a shorter wavelength, as seen here.



Ray tracing in *Mathematica*

Ripple simulation

Mercury drops in dish of mercury
Weber brothers, *Wellenlehre*, 1825

Figure 2.20

Three completely different approaches to the same phenomenon: an off-center point source of waves in a circular enclosure. The left image is a ray tracing, obtained by following rays from the off-center source point outward until they hit the circular walls, and then taking specular bounces. The middle image shows the *Ripple* simulation with a sinusoidal source. The right-hand image is the most remarkable, obtained in painstaking detail by watching the wave pattern from liquid mercury dropped periodically at the off-center source point in a circular dish of liquid mercury. The original drawing, published in 1826 in *Wellenlehre*, by Ernst Heinrich Weber and Wilhelm Eduard Weber, consists of about 200,000 individual handmade dots stylistically representing the antinodes as shaded-relief diamonds. Here, we clearly see the relation between two types of modeling (solving numerical equations simulating the waves as in *Ripple*, and following rays from the source) and the "real thing," as drawn in 1825 from an experiment.

Box 2.2

The SOFAR Channel

In the ocean, the speed of sound increases about 4 m/s for every 1°C increase in temperature. Variable temperature in the ocean therefore results in variable sound speed, which has the dual effect of refracting waves and making them arrive sooner or later according to the temperatures through which they have passed. In the 1970s, Walter Munk and Carl Wunsch suggested the idea of ocean acoustic tomography: measuring the temperature of the ocean remotely over large areas and at great depths if desired by sending sound long distances underwater to receiving hydrophones. Sounds are created hundreds of kilometers away from receiving stations, and ray tracing is used to help deduce the temperature profile of the water between the source and the receiver. Depending on the temperature profile, there may be many ways rays can travel from the sound source to the receiver, and generally each path will arrive at a different time and will have passed through different parts of the ocean. This technique is an important part of global earth monitoring of climate change.

Going deeper into the ocean, temperature steadily declines, but the pressure is rising. High pressure increases water density and causes an increase in sound speed. At first, the temperature decline wins, and sound speed decreases with depth. Eventually, at a depth of about 800 meters, the rising pressure overcomes the decline in temperature and causes the sound speed to go up again (see figure 2.21). Thus, as discovered in the United States toward the end of World War II and independently in 1946 in Russia (and kept top secret during the Cold War by both the American and Russian navies), there is a band of minimum sound speed, called the Sound Fixing and Ranging (SOFAR) channel, just under 1 km deep. Since waves always refract toward regions of lower speed, the SOFAR channel is a *waveguide* for sound. Communication over long distances is possible, since the sound energy is held to the channel rather than spreading in three dimensions. Figure 2.22 shows the results of two simple experiments in *Ripple*, with

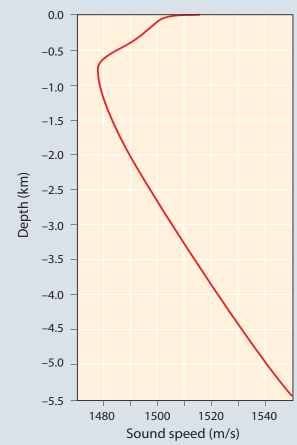


Figure 2.21

Speed of sound versus depth in the ocean. Courtesy Bdushaw.

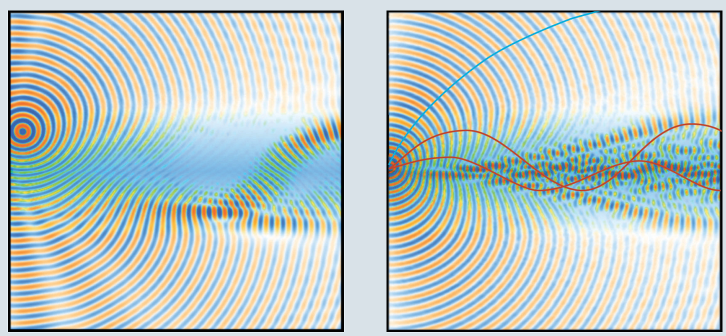


Figure 2.22

Simulations of sound-wave guiding in and near the SOFAR channel, conducted in the *Ripple* Java applet. (Left) Waves are launched from the left, above the middle of the channel (the channel is indicated as a blue hue; the sound speed reaches a minimum in the middle of this region), and pass through the channel. Some of the waves are refracted back and are trapped, oscillating from side to side as they progress down the channel. (Right) Some fraction of the waves launched inside the channel with low-enough angle are captured by the slow-speed channel. Once captured, the waves lose much less amplitude with distance traveled than they would outside the channel. Two representative trapped ray paths are shown in red; an escaping ray is shown in blue. Whales are thought to dive down nearly a kilometer in order to communicate using this sound conduit of the deep, which allows communication perhaps for thousands of kilometers.

The SOFAR Channel *(continued)*

the setting Temperature Gradient 4, which has a band of minimum sound speed, just like the SOFAR channel.

When ships lowered speakers and *hydrophones* (microphones designed to work well underwater) to use the channel, they found they were not alone: strange sounds are heard, believed to be coming from humpback whales diving down to take advantage of the waveguide effects. These whales may communicate with other humpbacks hundreds or perhaps thousands of kilometers away.

2.10 Measures of Sound Power

We have already had a few occasions to discuss sound power. We mentioned the enormous difference between the softest audible sound and sound at the threshold of pain. We have examined the falloff in sound intensity with distance coming from a small source. Whenever a quantity can vary by factors of millions or billions and yet be of significance over its whole range, we need to bring logarithms to the rescue. The *logarithm* measures the *exponent* that gives the number. The *base* of the logarithm is the number we are raising to a power, and the *power* is the logarithm of the number. Thus, by definition $7.3485 = 10^{\log_{10}(7.3485)}$. (We write the base as a subscript on the log.) By trial and error if need be (but of course we now have computers and before that, log tables), you can show that $7.3485 = 10^{0.866199}$, or $\log_{10}(7.3485) = 0.866199$.

Instead of talking about sound intensity, we use the log, which we also call intensity—that is, sound intensity measured in decibels (dB). The formula is simple:

$$I(\text{dB}) = 10 \log_{10}\left(\frac{I}{I_0}\right), \tag{2.5}$$

where $\log_{10}(\ldots)$ is the base 10 logarithm, and I_0 is a reference intensity, usually defined to be the threshold of human hearing. Thus, a 0 dB sound (the logarithm of 1 is 0) is barely audible to those with excellent hearing in a perfectly quiet environment. Since $10 \log_{10}(2) \approx 3$, a doubling of sound intensity (for example, two identical instruments instead of one) corresponds to a 3 dB increase. This definition makes clear that a 10 dB increase in sound intensity (measured in dB) corresponds to a ten-fold increase in power. The buzz of a nearby mosquito is about 40 dB, and a normal conversation is about 60 dB. There is 100 times more power in a normal conversational voice than in a mosquito buzz. Still, the mosquito is surprisingly loud, considering a human weighs 10 million times as much as a mosquito.

Other quantities are routinely given as their logarithms. The Richter scale for earthquakes comes to mind; it measures the base 10 log of the *amplitude* of motion of the earth.

Another reason for the decibel measure is that our hearing is essentially logarithmically sensitive. Our impression of loudness is not proportional to sound power, but rather approximately proportional to the *logarithm* of sound power. Table 2.1 gives some typical sounds and their corresponding power in decibels. Sustained exposure to 85 dB sound is considered harmful to hearing; needless to say, a rock concert at 110 dB is almost unquestionably going to do permanent damage.

In section 2.2, we discussed the falloff of sound power with distance from the source. We showed that the power passing through a window drops as $1/r^2$, where r is the distance from the source. There we also introduced the falloff in the *subjective* loudness, which is less rapid. If the distance doubles, from $r \to 2r$, the *objective* sound power I drops from $I = A/r^2$ to $A/4r^2$, where A is characteristic of the source. In terms of decibels, we have

$$I_{dB}(r) - I_{dB}(2r) = 10\log_{10}\left[\frac{I(r)}{I(2r)}\right] = 10\log_{10}\left(\frac{4}{1}\right) = 6.02,$$

that is, there is a 6 dB drop in sound intensity measured in decibels for every doubling of distance. This does not account for the effect of the ground, wind gradients, and the like.

Unless they are somehow maintaining a lockstep phase relation, the total intensity of two sources is simply the sum of the individual intensities. For two equally loud trumpets, intensity is $2I$ as compared to I for one trumpet, thus a 3 dB increase is seen when doubling the number of instruments.

Table 2.1

Some Sounds and Their Decibel Equivalents

Decibels	Pressure in pascals	Sound
0	0.00002	Threshold of human hearing
10	0.0000632	Human breathing at 3 m
20	0.0002	Rustling of leaves
40	0.002	Residential area at night
50	0.00632	Quiet home with some appliances on
70	0.0632	Busy traffic
80	0.2	Vacuum cleaner
90	0.632	Loud factory
100	2	Pneumatic hammer at 2 m
110	6.32	Accelerating motorcycle at 5 m
120	20	Rock concert
130	63.2	Threshold of pain
150	632	Jet engine at 30 m (hearing severely damaged)
180	20,000	Rocket engine at 30 m (near-instant death)

Ten trumpets are 10 dB louder than one trumpet, a tenfold increase in power. Because of our logarithmic hearing, ten trumpets subjectively sound only about twice as loud as one.

We are not equally sensitive to sound power at all frequencies—we will delay that discussion until chapter 22.

We can now return to the example at the start of this chapter. Is it reasonable that the sound of a bell can travel 100 miles and be heard aboard a ship? One hundred miles is about 160,000 meters, so the drop in decibels from, say 100 m distance from the town square to 160,000 meters is $10 \log[(100/160,000)^2] \approx 64$ dB. Suppose we say the bell 100 m away was a loud 95 dB. This becomes only about 30 dB at the ship, certainly masked by the probable 50 to 70 dB ambient noise aboard a ship. If we assume the sails can reflect 30% of the sound energy incident on them, and that they concentrate the sound by a factor of 1000 by focusing (see figure 2.7), we have amplification by a factor of 300 at the focal spot on deck.[2] This $10 \log[300] = 25$ dB, so that the sound at the focal spot would be $30 + 25 = 55$ dB; still very soft and probably inaudible except on a very quiet ship, quieter than most houses today.

However, we have been assuming uniform spreading of the sound according to the $1/r^2$ law. On most days, this is not at all the case for outdoor sound propagation over long distances. One reason is gradients in wind velocity, as we shall spell out in more detail in section 28.2. The wind is slowest at ground or sea level, being diminished by friction with the ground, and faster aloft. Sound traveling downwind, therefore, is slowest in its progress at ground level, and faster aloft, since it travels at 344 m/s through air. Sound or indeed any of the usual kinds of waves (light, water waves, and so on) refract toward regions of slower propagation. Therefore, if the wind was somewhat offshore, the downwind portion of the sound, traveling toward the ship, would have been spreading out not in three dimensions, but rather in two dimensions, being prevented from going aloft by refraction. There is only a 33 dB drop for a $1/r$ falloff of sound intensity, as opposed to the 65 dB for $1/r^2$.

In fact, the pattern of sound propagation and intensity downwind is not uniform for another reason, as has been chronicled many times in war and after accidental explosions. This subject is taken up in section 28.2. On a scale of 150 km or more from the source, sound propagation can be controlled by ever-present temperature gradients in the atmosphere, causing a refocusing of the sound that escaped aloft down to the ground about 150 km to 200 km from the source, making this sound many decibels louder than it would have been without the long-range refraction.

[2] Here, we are on the shakiest ground. Clearly, it is impossible for us to know precisely what the sail was doing that day. It is unlikely that it was a perfect shape for concentrating the sound. But decibels are a logarithmic measure, and significant errors of estimation end up as modest changes to the outcome.

Even without these atmospheric focusing effects, the bells might have been just barely audible at the focal point of the sail aboard a very quiet ship. Given more favorable atmospheric conditions, and favorable position and shape of the sail, there seems to be absolutely no doubt that the ringing of the church bells could have been heard 100 miles away.

Box 2.3

How Big?

It can be useful to calculate how big things are. It is easy to be wrong by many orders of magnitude when guessing how far a surface has to move in making a sound wave, or how much energy is in the sound wave, and so on. Grammy Award–winner Eberhard Sengpiel, sound engineer par excellence and lecturer at the Berlin University of the Arts, has compiled a very useful table (table 2.2) that makes it simple to calculate relevant acoustical quantities. Some of the quantities in this table we have not discussed specifically in this book, but we give them here nonetheless for reference.

Please note: The intensity I in the tables is *not* in dB, but bears the relation

$$I\ (\mathrm{dB}) = 10\ \log_{10}\left(\frac{I}{I_0}\right), \qquad (2.6)$$

where

$$I_0 = 10^{-12}\ W/\mathrm{m}^2.$$

As an example of the use of the table, suppose we want to discover how far a wall has to move (its displacement amplitude ξ) to create a very loud, 100 dB sound wave of frequency 1000 Hz. First, we find I from equation 2.6 by inserting 100 dB: $100 = 10\ \log_{10}\left[\frac{I}{10^{-12}}\right]$; or $I = 10^{-2}$. Then, from the table, we have

$$\xi = \frac{1}{\omega}\sqrt{\frac{I}{Z}} = \frac{1}{2000\pi}\sqrt{\frac{10^{-2}}{420}} \sim 8 \times 10^{-7}\ \mathrm{m},$$

just under 1 micron. The acceleration a is

$$a = \omega^2\xi = (2000\pi)^2 \times 8 \times 10^{-7} \sim 31\ \mathrm{m/s}^2,$$

or more than three times the acceleration due to gravity.

For a 100 Hz, 100 dB tone, the displacement is 10 times larger and the acceleration is one-tenth as large as for a 1000 Hz, 100 dB tone. If I is 40 dB at 1000 Hz, a soft sound but still well above the threshold for hearing in a home environment, $I = 10^{-8}\ \mathrm{W/m}^2$, a million times smaller than for 100 dB. The displacement ξ is proportional to square root of the intensity, so ξ is 1000 times less at 40 dB than at 100 dB or at 1000 Hz $\xi \sim 10^{-9}$m—just one nanometer, or the width of a few atoms!

How Big? *(continued)*

Table 2.2

Table of Formulas for Calculating Physical Quantities Related to Sound for a Traveling Plane Wave

	ξ	v	a	p	I	E	P_{ac}
ξ (particle displacement)	—	$\frac{v}{\omega}$	$\frac{a}{\omega^2}$	$\frac{p}{\omega \cdot Z}$	$\frac{1}{\omega}\sqrt{\frac{I}{Z}}$	$\frac{1}{\omega}\sqrt{\frac{E}{\rho}}$	$\frac{1}{\omega}\sqrt{\frac{P_{ac}}{Z \cdot A}}$
v (particle velocity)	$\xi\omega$	—	$\frac{a}{\omega}$	$\frac{p}{Z}$	$\sqrt{\frac{I}{Z}}$	$\sqrt{\frac{E}{\rho}}$	$\sqrt{\frac{P_{ac}}{Z \cdot A}}$
a (particle acceleration)	$\xi \cdot \omega^2$	$v \cdot \omega$	—	$\frac{p \cdot \omega}{Z}$	$\omega\sqrt{\frac{I}{Z}}$	$\omega\sqrt{\frac{E}{\rho}}$	$\omega\sqrt{\frac{P_{ac}}{Z \cdot A}}$
p (sound pressure)	$\xi \cdot \omega \cdot Z$	$v \cdot Z$	$\frac{a \cdot Z}{\omega}$	—	$\sqrt{I \cdot Z}$	$c\sqrt{\rho \cdot E}$	$\sqrt{\frac{P_{ac} \cdot Z}{A}}$
I (sound intensity)	$\xi^2 \cdot \omega^2 \cdot Z$	$v^2 \cdot Z$	$\frac{a^2 \cdot Z}{\omega^2}$	$\frac{p^2}{Z}$	—	$E \cdot c$	$\frac{P_{ac}}{A}$
E (sound energy density)	$\xi^2\ \omega^2 \cdot \rho$	$v^2 \cdot \rho$	$\frac{a^2 \cdot \rho}{\omega^2}$	$\frac{p^2}{Z \cdot c}$	$\frac{I}{c}$	—	$\frac{P_{ac}}{c \cdot A}$
P_{ac} (sound power)	$\xi^2\ \omega^2 \cdot Z \cdot A$	$v^2 \cdot Z \cdot A$	$\frac{a^2 \cdot Z \cdot A}{\omega^2}$	$\frac{p^2 \cdot A}{Z}$	$I \cdot A$	$E \cdot c \cdot A$	—

Angular frequency $\omega = 2\pi f$ Area A in m^2

Force in newtons $1\ \text{N} = 1\frac{\text{kg}\cdot\text{m}}{\text{s}^2}$ Density of air $\rho = 1.204\frac{\text{kg}}{\text{m}^3}$ at 20°C

Sound pressure ρ in pascals $1\frac{\text{N}}{\text{m}^2} = 1\ \text{Pa}$

$Z = \rho \cdot c$ = specific acoustic impendance of air $= 420\frac{\text{Pa}\cdot\text{s}}{\text{m}}$ at 20°C

Sound power P in watts $\left(1\ \text{watt} = 1\frac{\text{kg}\cdot\text{m}^3}{\text{s}^3} = 1\frac{N\cdot\text{m}}{\text{s}} = 1\ \text{joule/s}\right)$

Speed of sound $c = 343\frac{\text{m}}{\text{s}}$ at 20°C