Part E: Hearing & Psychoacoustics

Chapter 63. Intensity Just Noticeable Difference

In the early 1800s, the German physician Ernst Weber was one of the first people to study the science of perception. His work was not particularly focused on hearing, but one of his major contributions is a relation that applies to many different senses, including hearing. It involves the idea of a **just noticeable difference** (**jnd** for short, also called a **difference limen**), the degree by which an external stimulus needs to change in order to be just barely perceived as a change. Two stimuli that differ by less than one jnd would be perceived as being the same.

> Weber's Law: The just noticeable difference for increase in a stimulus is proportional to the initial stimulus.

For example, the louder a sound is, the greater intensity change is required to cause a perceptable difference.

As always, a proportion suggests taking a ratio. Another way to state Weber's Law, for the specific case of sound intensity, is with the equation

$$\frac{\Delta I_{\mathrm{jnd}}}{I} = K_I \quad , \tag{63.1}$$

where I is the intensity of the initial stimulus, ΔI_{jnd} is the intensity just noticeable difference (I-jnd for short), and K_I is the **Weber constant** (here, the subscript I indicates the stimulus is sound intensity). The Weber constant is different for different stimuli. For sound intensity, it is roughly

$$K_I = 0.25 \quad . \tag{63.2}$$

The big idea here is that the jnd depends on the initial intensity. Starting at $I_1 = 1 \times 10^{-4}\,\mathrm{W/m^2}$, the intensity would have to increase by $\Delta I_{\mathrm{jnd}} = K_I I_1 = 0.25 \times 10^{-4}\,\mathrm{W/m^2}$ to $I_2 = 1.25 \times 10^{-4}\,\mathrm{W/m^2}$ in order to be perceptably different. But starting at $I_1 = 1 \times 10^{-6}\,\mathrm{W/m^2}$, the required change is only $\Delta I_{\mathrm{jnd}} = 0.25 \times 10^{-6}\,\mathrm{W/m^2}$.

For a variety of reasons, Weber's Law cannot be expected to be extremely precise or universally true. There are variations between different people. Even for one person, stimuli pairs that are very near one jnd will be sometimes distinguished and other times not, so that the measurement of jnd becomes a statistical question. In addition, as might be expected for a complex biological system, there are deviations from the law, especially near the extremes of perception.

Nevertheless, taking a physics viewpoint (or more specifically, a **psychophysics** viewpoint), this is a nice simple model that adequately describes a wide variety of situations. Indeed, Weber's Law works particularly well for sound intensity, compared to other stimuli, being applicable over nearly the entire audible range.

In the mid 1800s, Gustav Fechner elaborated on Weber's observations. The law that now bears his name, introduced in Chapter 54, is in some sense just a restatement of Weber's Law.

> Fechner's Law: Human sensations are logarithmically related to the physical quantities that they sense.

Fechner's key addition is that "level" quantities such as sound intensity level, defined with a logarithm as in the equations of Chapter 57, are useful. Eq. 63.2 translates to an SIL-jnd of

$$\Delta L_{I,\mathrm{jnd}} = (10\ \mathrm{dB}) \log\left(\frac{I + \Delta I_{jnd}}{I}\right) = (10\ \mathrm{dB}) \log(1 + K_I) = 1\ \mathrm{dB} \quad . \tag{63.3}$$

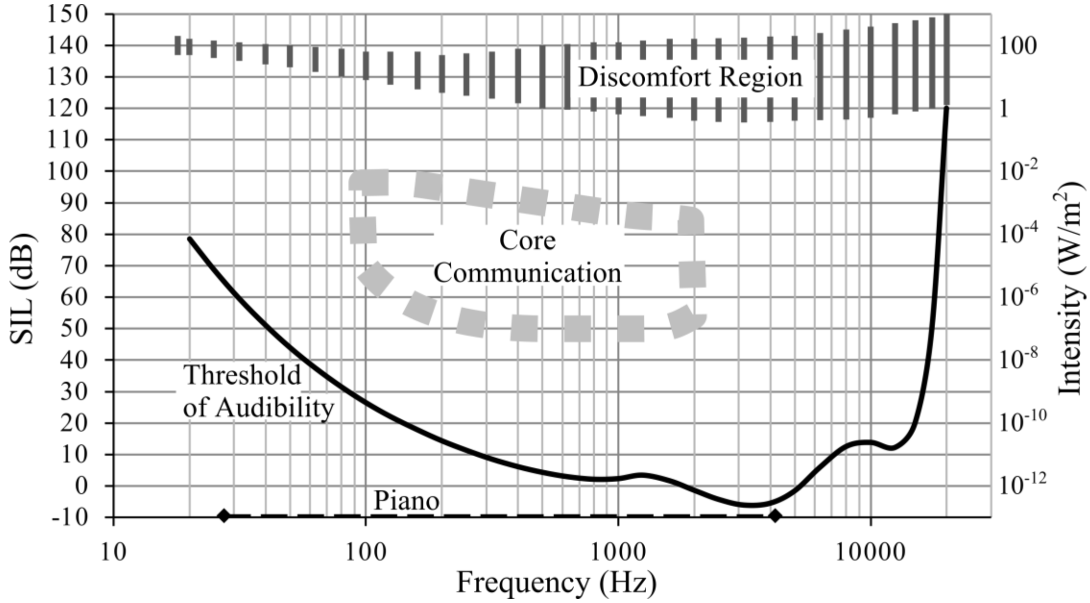

Figure 64.1

Range of human hearing. Discomfort range (vertical lines) roughly based on data from Licklider.[17] Threshold of audibility (solid line) according to the ISO 226:2003 standard,[15] extrapolated for 12.5 kHz to 20 kHz.

Therefore, the SIL jnd is always the same 1dB, independent of the initial SIL, and therefore easier to use. This is one of the reasons that the SIL scale is useful.

Despite having been established so long ago, increasingly careful measurement of the I-jnd in a variety of background conditions remains an active area of psychoacoustic research.[14]

Chapter 64. Hearing Thresholds

Roughly speaking, the human hearing range is 20 Hz to 20 kHz in frequency and 0 dB to 120 dB in SIL. But that description misses nuances, including a large set of low frequency, low SIL sounds that are inaudible. Figure 64.1 shows the details. Notice the logarithmic frequency axis, rather than the linear frequency axis that was used in many preceding chapters. This is another occurrence of Fechner's Law, further explored in Chapter 76.

Some features of Figure 64.1 could be described in terms of maximum and minimum frequencies. But the usual terminology treats frequency as the independent variable. The various curves and limits, introduced in Chapter 51 as single numerical quantities, are more accurately described as intensities or SILs which are functions of frequency. Partly, this choice may be due to perception. Perceived pitch is determined almost entirely by frequency, but perceived loudness depends on more than intensity, as detailed in Chapter 66.

None of the boundaries in Figure 64.1 are really as distinct as the figure might suggest. If a sound changes frequency or loudness so that it approaches a boundary, the sound will smoothly transition into a different sensation, or no sensation at all. On top of that, there is variation between individuals, for a variety of reasons. Even though there are firm standard numbers or mathematical functions for some of the limits, keep in mind that they are phenomenological and situational.

---

[14] Jennifer H. Johnson, Christopher W. Turner, Jozef J. Zwislocki, and Robert H. Margolis, "Just noticeable differences for intensity and their relation to loudness," *J. Acoust. Soc. Am.* 93(2) (1993): 983-991.

The most clearly defined curve is the threshold of audibility, the solid line in Figure 64.1. For example, the curve dips towards lower SIL around 3500 Hz. Human ears are more sensitive near that frequency, so less intensity is required to be perceptable. As frequency decreases below 600 Hz, human ears are steadily less sensitive, so that a higher and higher SIL is required for a sound to be perceived. Of course, any one individual may or may not perceive any specific sound near the boundaries, there is variability among different individuals, and there is also a judgment as to what constitutes "normal" hearing. Nevertheless, much careful work has gone into this area, and the threshold of audibility in Figure 64.1 is based on conclusions from the International Standardization Organization, with some extrapolation at the very highest frequencies.[15]

At high intensities are the thresholds for discomfort, a tickling feeling, and pain. It is quite difficult to identify specific values for these threshold intensities. Not only can the SILs vary by 10 dB or more for different individuals and different timbres, but also habituation (due to lengthy exposure) can raise the threshold by another 10 dB or so.[16] The vertical lines in Figure 64.1 roughly show the region where sound is uncomfortable, but not painful, based on data compiled by Licklider.[17]

At the low frequency side, the ear does not simply become insensitive to sound. Instead, below 20 Hz the sensation transitions from a tone to a rapid popping or pulsing before becoming imperceptible. It's not possible to identify a precise frequency at which this happens, but it is essentially the same for all audible intensities.

For comparison, the frequency range of the pitches on a piano is also shown (dashed line). With very few exceptions, the piano range encompasses all of the pitches used on any other musical instruments. However, keep in mind that sounds, including musical tones, get their timbre from overtones at higher frequencies than the fundamental. The core communication region in Figure 64.1 includes the frequencies for the fundamental and several overtones of the human voice (both genders), and roughly the intensities that might be encountered in speaking or singing.

Chapter 65. Hearing Level

Probably most people in the United States have had a basic hearing test at some point, in which progressively fainter tones at a variety of frequencies are presented to each ear. The result of such a test is usually presented in an **audiogram** such as Figure 65.1, which shows the subject's hearing threshold as a function of frequency, once again plotted on a logarithmic axis. But, if the human audibility threshold increases markedly below 600 Hz, how can this audiogram show results near 0 dB at the low frequencies? And why does the vertical axis increase downwards?

The key is that the vertical axis is not measuring the threshold as an SIL, but instead as a **hearing level** (abbreviated **HL**). This scale uses the same units as SIL, but it sets the 0 dB level to be equal to the normal threshold of audibility. Thus, hearing level indicates the extra intensity, *above the normal threshold*, needed for this individual to hear a tone. This relationship is expressed by the equation

$$L_{I,\mathrm{min}} = L_{I0} + HL \quad , \tag{65.1}$$

where $L_{I,\mathrm{min}}$ is the individual's threshold intensity level, and L_{I0} is the normal threshold intensity level (the solid curve in Figure 64.1). Needing extra intensity to hear a sound is a bad thing, so the scale is inverted

[15] International Standardization Organization, *ISO 226: 2003(E): Acoustics—Normal Equal-Loudness-Level Contours* (Geneva: ISO, 2003).
[16] Edwin B. Newman. "Speech and Hearing," in *American Institute of Physics Handbook*, 3rd ed., ed. Dwight E. Gray (New York: Mcgraw-Hill, 1972), 3-155.
[17] J. C. R. Licklider. "Basic Correlates of the Auditory Stimulus," in *Handbook of Experimental Psychology*, ed. Stanley S. Stevens (New York: John Wiley & Sons, 1951), 995.

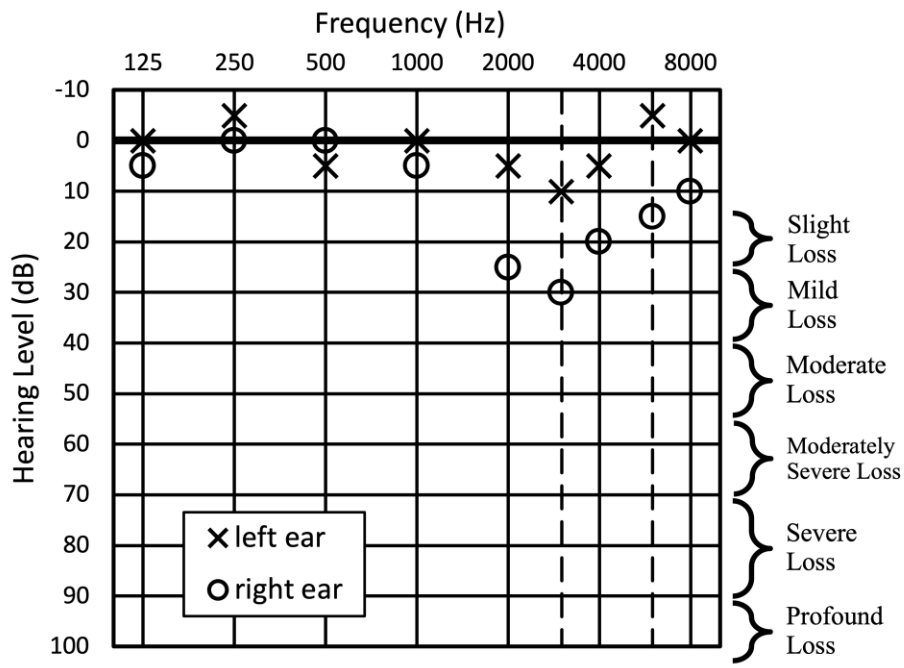

Figure 65.1

Typical audiogram, showing mild hearing loss in the right ear near 3000 Hz

(since we are used to associating "down" with "bad"). Figure 65.1 indicates on the right some qualitative descriptions of severity for various hearing level ranges.[18]

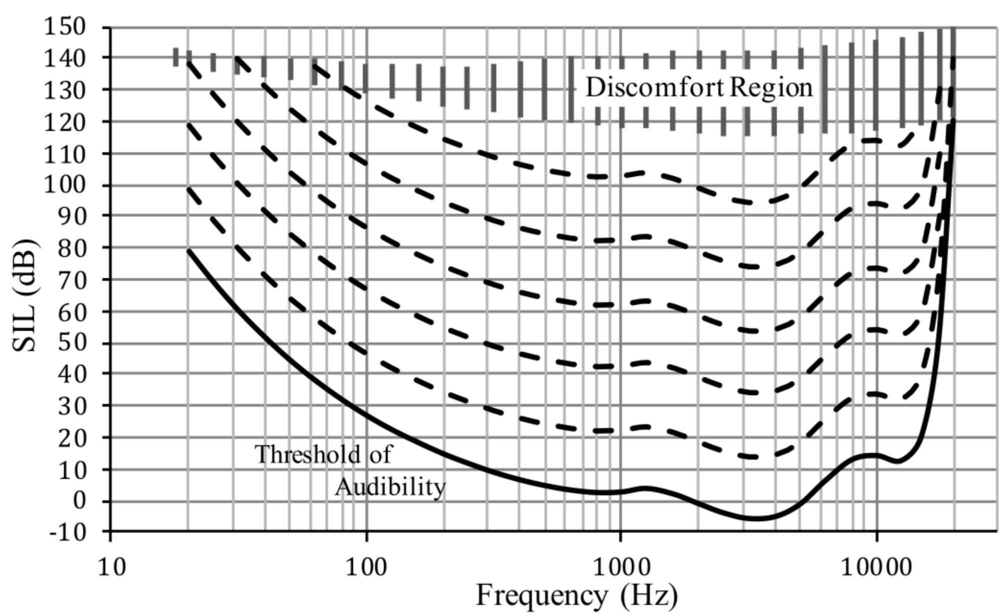

Figure 65.2

Lines of constant hearing level: Dashed lines are duplicates of audibility threshold, shifted upwards by multiples of 20 dB.

[18] J. G. Clark, "Uses and abuses of hearing loss classification," *Asha* 23(7) (1981): 493–500.

Because the decibel unit is used for both SIL and HL (and, indeed, many other quantities), the nature of a measurement is often indicated by adding to the unit, as in 30 dB SIL or 5 dB HL. These are not truly different units, however, so it is acceptable that they are being added together in Eq. 65.1. You may be familiar with the analogous situation in measuring temperature in units of kelvin (K) or degrees Celsius (°C); a temperature change of 1 K is the same as a change of 1°C, but the two units have different zero levels. The units dB SIL and dB HL are similar, with the added twist that the zero level of dB HL depends on frequency. Another analogy is the difference between altitude above sea level and height above ground level, where the zero level for height depends on your geographic location.

In Figure 65.2, the dashed curves are lines of constant hearing level, meaning that they are copies of the audibility threshold that are shifted upwards. If a person had a hearing level that was the same for all frequencies, then these curves would represent their personnel SIL audibility threshold. That's a very unlikely scenario, but it is one way to understand what hearing level is. Keep in mind that hearing level is used to describe a person, not a sound. A connection between hearing level and sound intensity level can only be made with reference to a specific individual.

Chapter 66. Perceived Loudness

Not only does the threshold of audibility depend on frequency, but the apparent increase in loudness with each dB SIL is also variable. In 1933, Harvey Fletcher and Wilden Munson addressed this by measuring what SIL was required for their subjects to perceive sounds as having equivalent loudness.[19] The results, updated to reflect research since that time, are shown in Figure 66.1. This sort of graph is sometimes called a **Fletcher-Munson diagram.**

Often this book uses the word loudness to refer to the qualitative sensation, but this graph defines **loudness level** quantitatively, measured in units of phons. Loudness level is defined such that for a pure tone at the

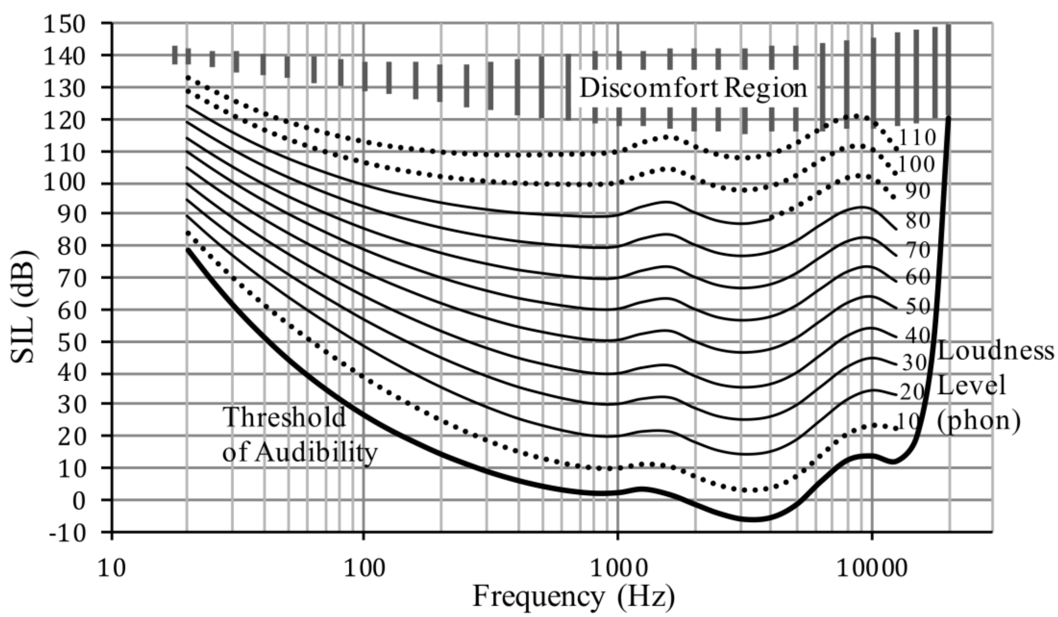

Figure 66.1

Curves of equal apparent loudness, defining the phon unit. Loudness level curves according to ISO 226:2003 standard.[15]
Dotted lines extrapolate standard outside its stated range of applicability.

[19] Harvey Fletcher and Wilden Munson, "Loudness, its Definition, Measurement and Calculation." *J. Acoust. Soc. Am.* 5 (1933): 82–108.

reference frequency of 1 kHz, loudness level in phons numerically matches the sound intensity level in dB. At other frequencies, the loudness level is the same as the level of a sound at 1 kHz that is perceived as having the same loudness.

A key feature of these curves of equal loudness is that they get closer together at low frequencies. An increase of 1 dB SIL can result in a loudness increase of as much as 2 phon. The effect is most prominent below about 300 Hz, although above 10 kHz it also occurs to a small extent. This squeezing together of the curves is the primary difference between Figure 66.1 and Figure 65.2, if you read that chapter.

This has a direct implication for adjusting the volume on audio systems. Suppose that you have your stereo system set so that the sound out of the speakers is just the way you want it, both in loudness and in the balance between bass, mid-frequencies, and treble. To increase the volume, the simplest thing for the electronics to do is to multiply the intensity by the same factor at all frequencies. This means that the same dB SIL will be added at all frequencies. Because of the shape of the equal loudness curves, this means that the *perceived* change in the bass and high treble will be greater than the perceived change in the mid-range.

Home audio systems of moderate quality often have a button labeled "Loudness Compensation," or simply "Loudness," which is intended to correct for the problem described above. The concept is that the bass & treble settings are to be adjusted as desired at moderate volume with loudness compensation off. If the volume is then decreased, the high treble and especially the bass will become too weak. Turning on loudness compensation should provide the appropriate boost to extreme frequencies in order to maintain the same timbre.

Ideally the degree of this compensation should depend on how much the volume is changed from its original value. However, very often audio systems don't do this, and instead the "loudness" button simply boosts the bass (and possibly the high treble). To be used as intended, the loudness compensation would only be turned on at low volume levels. However, listeners who like strong bass in their music sometimes use the "loudness" button as a bass-boost button even at moderate or high volume levels, unaware of its intended use.

Chapter 67. Loudness Approximated

Sound level meters are calibrated microphones that can be used to measure how loud sounds are. The result is in units of decibel, but the measurements are not necessarily sound intensity level (SIL). This is

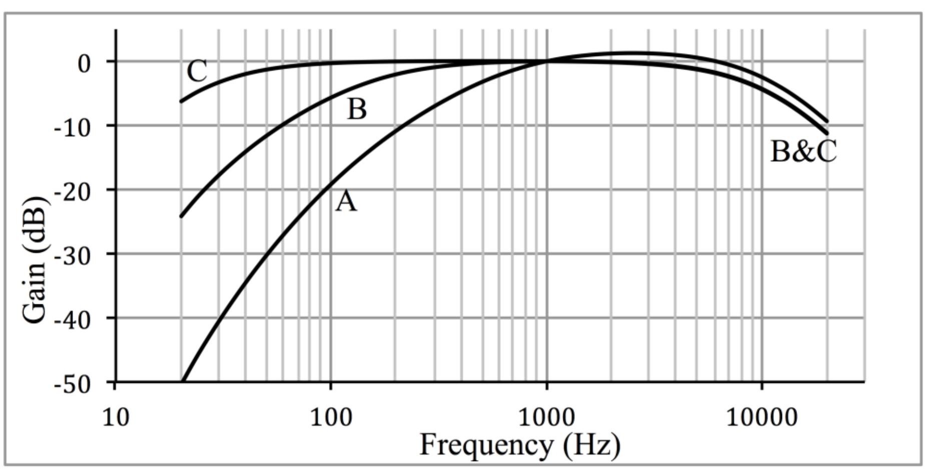

Figure 67.1

Weightings used by sound level meters. Gain is difference of meter response from physical SIL.

because sound level meters usually apply a **weighting,** which affects how much weight, or impact, sounds at different frequencies have on the measurement.

Several standardized weightings, as defined by the International Electrotechnical Commission, are shown in Figure 67.1.[20] Because these weightings compare different frequencies relative to each other, it is necessary to choose one reference frequency to which all others will be compared. The usual frequency to choose in such situations is 1 kHz, a nice round number that is fairly close to the center of the audible range. This is why all the weighting curves pass through 0 dB at 1 kHz. At other frequencies, the curves indicate how an actual, physical SIL measurement should be adjusted to get the weighted value. Negative gain means that the weighted value will be lower than the physical measurement.

The most important is the A-weighting, which is intended to roughly account for the fact that human hearing is less sensitive to low frequencies. The B-weighting is officially outdated, and it is rarely seen. There is also a D-weighting, which is only used in a very specialized situation.

Consider applying the A-weighting to a sound spectrum, to calculate the sound's A-weighted intensity level L_{IA}. Specifically, consider a spectrum with several partials, each one a well-defined peak with some SIL value. For each partial, we would add to its SIL the A-weighting gain appropriate for its frequency, to get that partial's A-weighted intensity level. The weighted intensity level of the whole sound is the combination of the levels of the partials. Keep in mind, however, that "combining" decibel measurements does not mean simply adding them. To calculate the total L_{IA}, one would need to find the weighted, or effective, intensity of each partial with Eq. 57.2, add those intensities as indicated by Chapter 60, and then convert back to dB with Eq. 57.1.

To indicate that weighted measurements are not the same as SIL, the unit is modified. For instance, the A-weighting unit is dB(A), sometimes written dBA or dBa.

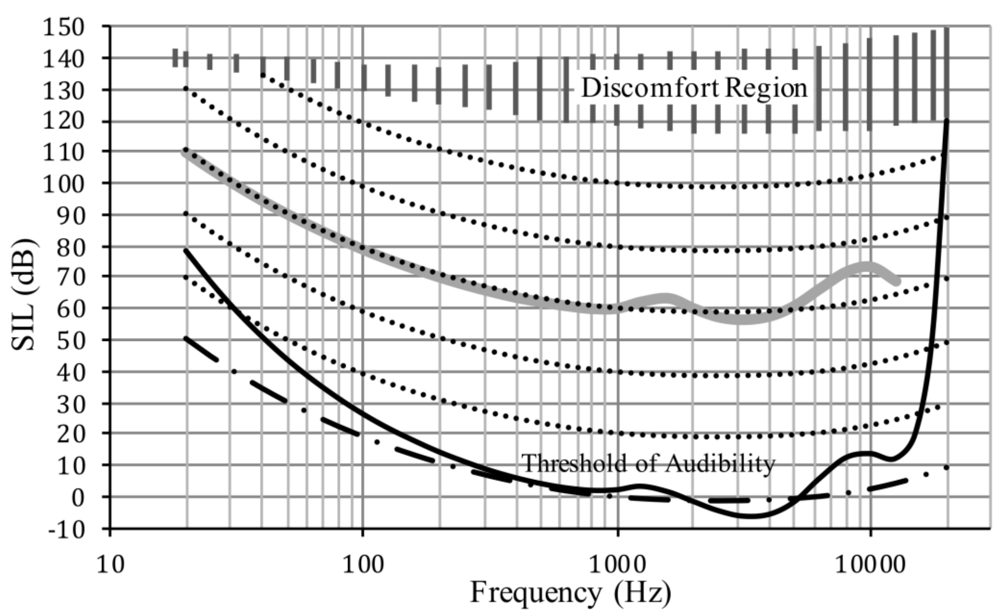

Figure 67.2

Dash-dot line: A-weighting approximation of the audibility threshold. Dotted lines: This threshold shifted upwards by multiples of 20 dB. Gray band: 60 phon level.

[20] International Electrotechnical Commission, *IEC 61672 Ed. 2.0 b:2013: Electroacoustics – Sound Level Meters* (Geneva: IEC, 2013)

One way to think of weighting is that it shifts the meaning of 0 dB as a function of frequency. (This is somewhat like hearing level, from Chapter 65.) In Figure 67.2, the dash-dot curve shows the SIL at which a pure tone would have an A-weighted sound intensity level of $L_{IA} = 0$ dB(A). That curve is an inverted version of the A-weighting shown in Figure 67.1. The other dotted lines serve the same function for higher values of L_{IA}; they are simply copies of the dash-dot curve, shifted to higher SIL.

Comparing the $L_{IA} = 0$ dB(A) curve in Figure 67.2 with the threshold of audibility, one might think that the A-weighting does not do a very good job of mimicking the response of the human ear, especially below 200 Hz. However, the 60 dB(A) curve is quite a good match to the thick gray curve, the 60 phon loudness level, especially below 5 kHz. The A-weighting is chosen as a best match to perceived loudness for moderate intensity sounds. Because the equal-dB(A) curves do not compress at low frequencies, the way the equal-phon curves do, the A-weighting overestimates loudness for lower-intensity low-frequency sounds, and underestimates loudness for high-intensity low-frequency sounds.

The C-weighting, which is commonly available, makes almost no adjustment for all frequencies in the human hearing range (although it has some decrease in gain at very low and high frequencies). Thus, results from a sound level meter on C-weighting come close measuring the physical sound intensity level. Some sound level meters have an un-weighted setting (sometimes called Z-weighting), which even more accurately measures physical SIL.

Chapter 68. Resonance

When a varying force causes an object to vibrate, the object and its vibration are both said to be **driven.** For instance, the vibration of an eardrum is driven by the varying force of the air on it. An ideal **non-resonant vibrator** is something that, if driven at many different frequencies, responds with vibration equally well at all frequencies. But most driven vibrators are not like this. Instead they respond more at some frequencies than at others.

For instance, Chapter 27 discusses how a mass on a spring has a specific natural frequency, with which it will oscillate when initially disturbed, but then left to itself. This is non-driven oscillation. Given this, it will perhaps not seem surprising that if a mass on a spring is driven (that is, by a varying force *in addition to* the spring), it will respond with a larger oscillation when the driving frequency matches the natural frequency. It is as if the mass-spring system particularly likes to oscillate at that frequency, and thus responds more to that frequency than any other.

To be more specific, imagine an object, with some specific mass, hanging from a spring, which in turn is hanging from your hand. (The following description will be far more believable if you actually try the experiment yourself.) Use the variable f_0 to mean the natural frequency of the object-and-spring. If you move your hand up and down sinusoidally at a frequency f and with amplitude x_{m0}, then the object will also move up and down. Figure 68.1 shows an example of the results for an experiment where x_{m0} is kept constant, but f is varied.

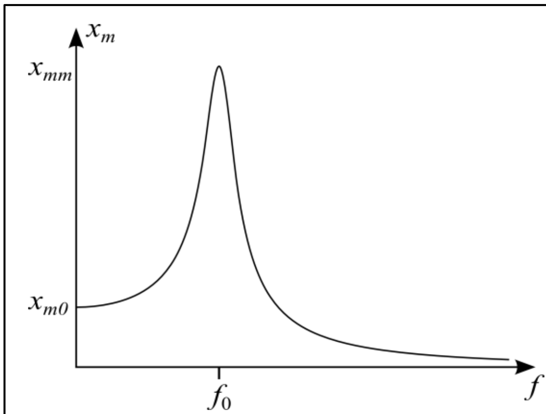

Figure 68.1

Oscillation amplitude of mass on spring driven by sinusoidal force with frequency f. f_0 is system natural frequency and $Q = 5$.

The first key observation is that the object will oscillate sinusoidally at the same frequency as your hand. The frequency f completely overrules any tendency for the object to move with frequency f_0. This may not seem to be true when you first start shaking your hand. The initial motion may even be rather erratic. But if you continue to drive the object at frequency f, it will eventually fall into an oscillation with frequency f. (How long you have to wait is related to the damping in the system, discussed in Chapter 33.)

Driving the oscillation at different frequencies will result in different amplitudes of the object. If your hand motion is very slow (small f), then the spring will stay close to its equilibrium length, so that the object moves with an amplitude equal to your hand's amplitude. But as the driving frequency increases closer to the natural frequency f_0, the spring will stretch and compress in time with your hand, so that the mass moves with a much larger amplitude. When your hand frequency matches the natural frequency, $f = f_0$, the object's motion will obtain the largest amplitude in the experiment, $x_{mm} \gg x_{m0}$. Changing to driving frequencies higher than f_0, the object will move with smaller amplitude, until at very high driving frequencies the object hardly moves at all.

A striking thing about this experiment is the large amplitude that occurs when the driving frequency is close to the system's natural frequency. This behavior is called **resonance.** Resonance is a behavior, not a quantity that can be measured. When the driver (your hand in the experiment above) has such a frequency, it is said to be **in resonance** with the oscillator.

The frequency at which the maximum response occurs is a characteristic of the system called its **resonant frequency.** (In the experiment above, the system includes the object and the spring, but not the moving hand.) A close mathematical analysis reveals that the resonant frequency is slightly lower than the natural frequency. However, the difference is very small, and often overlooked even in technical treatments, so that natural frequency and resonant frequency are considered synonyms. The choice of which term to use is determined more by context. When there is no driving force, f_0 would probably be called the natural frequency. But when there is a driving force, f_0 is more likely to be called the resonant frequency.

In a different experiment with the same equipment, you could keep the driving frequency at resonance, $f = f_0$, and vary the driving amplitude x_{m0}. As long as the spring is not deformed beyond where Hooke's Law is valid, in any real system the maximum amplitude turns out to be proportional to the driving amplitude, and their ratio

$$\frac{x_{mm}}{x_{m0}} = Q \tag{68.1}$$

is called the **quality factor** of the system.[21] A "real" system is specified here because the experiment only really makes sense if modeled including a third force. In addition to a restoring force and a driving force, there needs to be a damping force which tends to stop the system from oscillating. A large quality factor is associated with having a weak damping force, and vice versa. Chapters 33 and 34 go into more detail, but for some purposes we can describe a system with its quality factor, and otherwise ignore the damping force.

When the quality factor is very large, a tiny driving amplitude might be enough to get a rather large response. Imagine a version of Figure 68.1 in which the peak is very tall and very narrow. In that case, you might say that the oscillator hardly moves at all except when the driving frequency nearly matches the natural frequency. That is a pretty crude model of resonance, but for sometimes it's all that's needed.

Chapter 69. Complex Resonance

69a. Complex Driving Forces

Chapter 68 describes resonance when the driving force has a specific frequency, that is, it is simple harmonic. What if the driving force is more complicated, for instance arising from a complex vibration? Perhaps imagine a complex sound traveling through the air, and applying a force on an object that it reaches.

---

[21] For the purist: The quality factor is actually very slightly smaller than this. It is defined by a different formula, which this book will not cover. Like the difference between resonant and natural frequencies, in most cases this distinction is small enough to be ignored.

For non-resonant vibrators in Chapter 37, multiple different simple harmonic driving forces combined by superposition to make the vibrator move. Chapter 42 showed that a complex vibration is really the same as a combination, or superposition, of simple harmonic partials. A complex driving force is really no different from several harmonic driving forces working simultaneously.

A sympathetic vibrator with a resonant frequency really responds in a very similar way to a non-resonant one, but the details are a bit more complicated. Here is a rephrasing of the rule from Chapter 37, using some of the more technical terms.

> If a vibrator (either resonant or non-resonant) has linear response, then it obeys superposition. That is, its displacement in response to a complex driving force is the moment-by-moment sum of all the responses that it would have to the sinusoidal partials in the complex driving force.

Starting with the complex driving force, Fourier analysis can determine its partials. Each of those sinusoidal partials would result in some sinusoidal response of the same frequency, as described in Chapter 68. Those responses can then be recombined with superposition, to find the final complex response.

For an ideal non-resonant vibrator, the response is the same for all frequencies, so the final complex response will look just like the initial complex driver. The new element with a resonating vibrator is that the response might be different at different frequencies, changing the shape of the final complex response.

69b. Extra: Sympathetic Vibration

In some situations, a vibrating system is driven by a force that comes from another vibrating system. For example, a window rattles due to a loud radio, or a wine glass vibrates when a soprano sings. (Maybe the wine glass even shatters!) In general, when a sending vibrator A causes vibration in receiving system B, system B is called a **sympathetic vibrator,** and its motion is called a **sympathetic vibration.** The connection between them, technically referred to as the **coupling,** needs to be relatively weak, because if they were strongly connected then it wouldn't make sense to consider them as two separate oscillators. In the two examples given, sound is what couples the two oscillating systems.

Some references define the term sympathetic vibration to require that the frequency of B's vibration must match the frequency of A's vibration. But this is either pointless or misleading. If vibration A is a pure tone, as in the experiment described in Chapter 68, then the response is always at the same frequency as the driver. So, the restriction of matching frequencies is pointless, because it's the only possibility. On the other hand, if vibration A is complex, then it is entirely possible for an overtone of A's spectrum to be in resonance with a natural frequency of vibrator B. In that case, the frequency of the most prominent partial in B's spectrum might *not* match the frequency of the most prominent partial in A's spectrum. Nevertheless, that is still sympathetic vibration.

Sympathetic vibration is most obvious when the receiver B has a natural vibration frequency, and vibration A (or a partial of A) is in resonance with it. This is **sympathetic resonance,** which can lead to a much larger amplitude in B's vibration than when A is not in resonance with B. A singer shattering a wine glass is a dramatic example of this.

Some references fail to distinguish between sympathetic vibration and sympathetic resonance. But there are many examples of non-resonant sympathetic vibration. Eardrums are an example of sympathetic vibration without a resonance frequency, and the rattling window example may or may not involve a resonance.

Chapter 70. Ear Structure

Figure 70.1 shows the basic anatomy of the human ear. Terrestrial mammals all have a similar structure. The **outer ear** consists of those parts which are directly exposed to air outside of the body. The **auricle** is

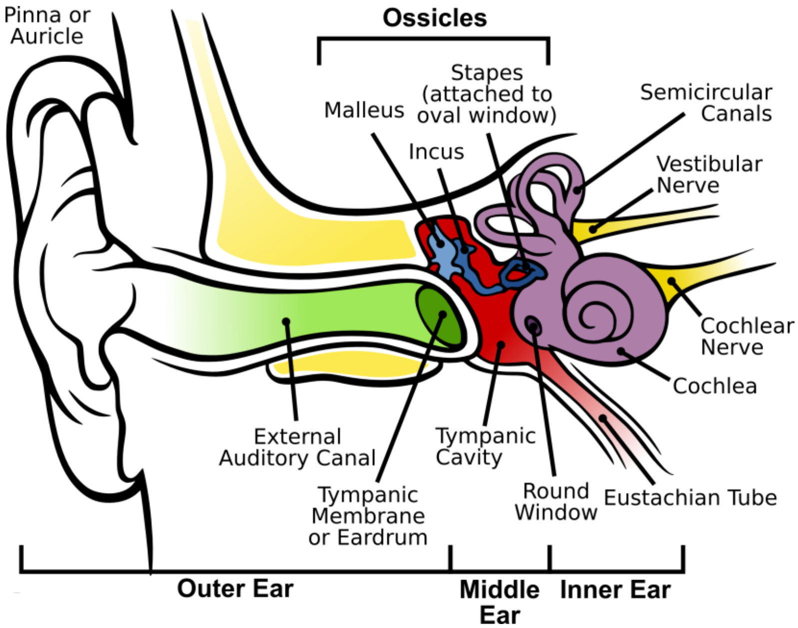

Figure 70.1

Anatomy of human ear. Reprinted with additional labels from Daniel Mietchen, "Anatomy of the Human Ear," *Wikimedian in Residence,* 2011, http://wir.okfn.org/2011/12/30/anatomy-of-the-human-ear-editable-version/ (October 2013) / CC BY 2.5.

the external part of the ear. This is also called the **pinna**, especially in the case of animal ears. It helps gather sound and funnel it into the **auditory canal,** more so in many other animals than in humans. This effect also helps in the determination of the direction from which sounds come. The auditory canal (and outer ear) terminates at the **tympanic membrane,** or **eardrum,** where the disturbance in the air is converted into the vibration of an object.

The eardrum is a non-resonant sympathetic vibrator, which means that it responds to a wide range of frequencies. Its name includes "drum" because it is a round membrane that is very thin: about 50 μm thickness, although data is surprisingly scant.[22] However, its shape is conical rather than flat, pointing inwards with an apex angle of about 120°, so it doesn't look much like the percussion instrument that it is named after.

The **middle ear** is also filled with air, delivered via the **Eustachian tubes** from the back of the pharynx in the throat. This air connection is important because the external air pressure (see Chapter 125) can change with the weather, altitude, and other conditions. An imbalance of pressure across the eardrum can be painful, and in extreme cases can even rupture the eardrum. However, the Eustachian tube is not always open. Opening the Eustachian tube to equalize the middle ear air pressure, which can be done voluntarily, results in a "popping" in the ear.

The work of the middle ear is done by the **ossicles,** three tiny bones which transmit the vibration of the eardrum to a vibration of the **oval window** on the liquid-filled **cochlea.** The names of the ossicles are Latin for shapes they vaguely resemble: a hammer (malleus), an anvil (incus), and a stirrup (stapes). One tip of the malleus is attached to the apex of the eardrum, and the flat side of the stapes covers the oval window. The arrangement of the ossicles creates a mechanical connection that efficiently transmits air vibrations into vibrations of the cochlear liquid, but the details of those mechanics will not be described in this book.

[22] W. Robert J. Funnell and Charles A. Laszlo, "A Critical Review of Experimental Observations on Eardrum Structure and Function." *ORL J. Otorhinolaryngol. Relat. Spec.* 44 (1982):184.

Chapter 71. Place Theory of Pitch

71a. Cochlear Strucure

The **inner ear** (see Figure 70.1), comprising the **cochlea** and **semicircular canals**, is a cavity in the bone of the skull, lined with and crossed by membranes. The semicircular canals are unrelated to the hearing function, instead providing our sense of balance. But the cochlea is the primary sound analysis organ.

Figure 70.1(b) shows the cochlea as it is physically shaped. The image is extracted from Figure 70.1, but it has been rotated so that the vertical axis of the head is down and to the right. (In this orientation, the semicircular canals would be to the lower right of the oval window.) The shape of the cochlea resembles a snail shell, except that as it winds towards the center it also moves towards the top of Figure 70.1(b), forming an overall conical shape. The auditory nerve enters the cochlea inside the cone, from behind the page in Figure 70.1(b). In humans, the coil wraps around 2¾ times.

The internal structure of the cochlea is easier to understand if we first imagine uncoiling it. Figure 70.1(c) shows the result of holding the windows in Figure 70.1(b) stationary and uncoiling the cochlea to the left. In the spirit of physics, Figure 70.1(c) is a **schematic,** which shows only the most essential elements in simplified geometric shapes. An anatomy book would show greater detail with more accuracy, but in any case, a real cochlea could not actually be uncoiled like this.

The human uncoiled cochlea would be about 3 cm to 3.5 cm long and about 3 mm wide at the end with the windows. It is somewhat tapered into a cone shape with a rounded apex, the base of the cone being the end with the windows. The interior is mostly a liquid-filled cavity. The **oval window** and the **round window** are covered with thin membranes, keeping the liquid in.

Along its length, from the wall that was originally towards the center of the spiral, the **bony shelf** or **osseous spiral lamina** (dark gray in Figure 70.1(c)) sticks out nearly half the width of the interior. From the opposite wall similarly extends the less prominent **spiral ligament,** as well as a secondary bony shelf near the base end**.** Between these is stretched the functional core of the cochlea, the **basilar membrane** (light gray in Figure 70.1(c)). Together, these three parts separate the cochlea interior into two chambers, which are connected only at the **helicotrema,** a small hole near the apex.

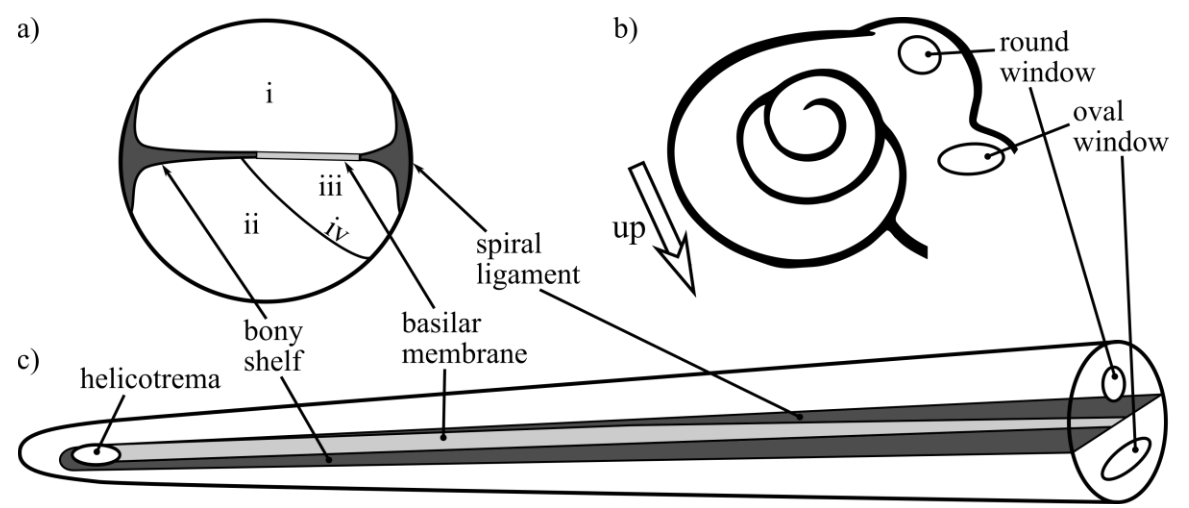

Figure 70.1

Structure of cochlea. (b) Physical shape of cochlea, rotated so that indicated direction is towards top of head. (c) Schematic of the cochlea uncoiled and showing some of basic internal structure; side towards viewer is side facing center of coil. (a) Schematic of a cross section of the cochlea.

A key feature of the basilar membrane is that it gradually changes from one end to the other. It tapers in the opposite direction from the cochlea, being about 0.1 mm wide at the cochlear base, and about 0.5 mm wide near the helicotrema. The basilar membrane is also thinner and more tightly stretched near the base, and conversely thicker and more slack near the helicotrema.

Attached to the basilar membrane, on the side closer to the oval window (the bottom side in Figure 70.1(c)), are the so-called **hair cells.** These cells do not have actual hair, but instead a few dozen protrusions at one end of the cell called **stereocilia.**

71b. Cochlear Function

When a sound enters the ear, parts in the **middle ear** vibrate against the oval window from outside of the cochlea. This vibration launches a sound wave, similar to sound in air, that travels through the cochlear liquid towards the helicotrema. As that liquid sound wave travels along the basilar membrane, it applies oscillating forces to the membrane, which try to drive the membrane into vibration. If the membrane vibrates, then it shakes the attached hair cells. The resulting motion of the stereocilia then generates nerve impulses that are sent to the brain.

The special feature of the cochlea is that for an incoming pure tone, the basilar membrane primarily vibrates at one place along its length. The location of that vibration depends on the frequency of the pure tone. This is the basis for the **place theory of hearing**, which says that we perceive different pitches based on the place that they make vibrate on the basilar membrane. Although this idea was investigated by Helmholtz[23] and Corti in the mid-1800s, the work clearly establishing this theory was done by Georg von Békésy starting in the 1930s.[24] For this work Békésy received the 1961 Nobel Prize in Physiology or Medicine.

Even though this is a rather complex system, some principles from Chapter 27 can qualitatively explain the basics. In that chapter we had the equation

$$f_0 = \frac{1}{2\pi}\sqrt{\frac{k}{m}} \tag{71.1}$$

for the natural frequency of a mass on a spring. The basilar membrane is much more complicated than a mass on a spring, but still it has a restoring force (from the tension pulling it tight) and it has a mass (depending on its thickness and width). The frequency with which different pieces of the basilar membrane will naturally vibrate depends on those characteristics in the same way as Eq. 71.1, with a stronger restoring force (like a higher spring constant) leading to a higher natural frequency, and a larger mass leading to a lower natural frequency.

Both of those characteristics vary along the length of the basilar membrane. Near the base, the thin, narrow membrane has less mass, and the membrane is more tightly stretched, so that the natural frequency will be quite high. Near the apex, the thick, wide membrane had more mass, and the membrane is more slack, meaning that there is a smaller restoring force. Thus, the natural frequency will be lower there.

When a complex sound enters the cochlea, keep in mind that it is a combination of pure tone partials. At each place where the natural frequency of the basilar membrane matches one of those partials, resonance causes a large amplitude vibration. Thus, through its mechanical structure, the basilar membrane essentially does Fourier analysis of the complex sound! The frequency axis is along the uncoiled cochlea, and the amplitude is indicated by the amplitude of basilar membrane vibration.

- 23 Helmholtz, *On the Sensation of Tone,* 49.

[24] Georg von Békésy, "Some Biophysical Experiments from Fifty Years Ago." *Annu. Rev. Physiol.* 36 (1974): 10.

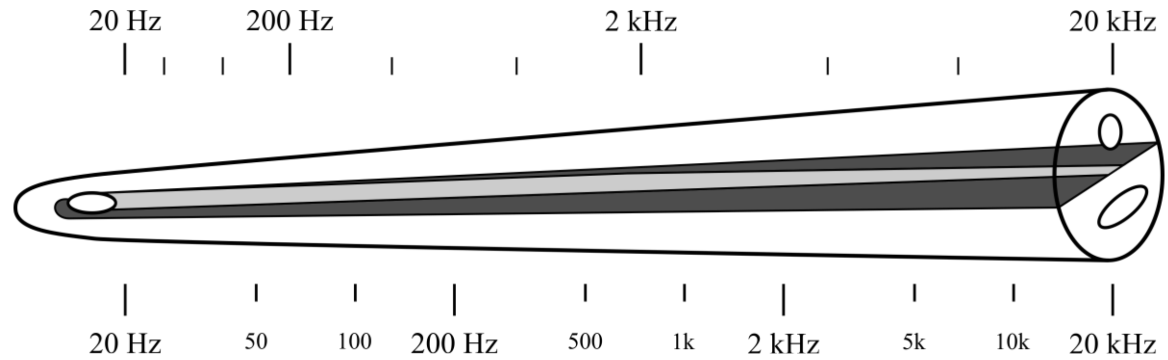

Figure 71.2

Position of resonant frequencies on basilar membrane. Top scale shows true positions based on data from Greenwood.[25] Bottom scale is idealized logarithmic model, as suggested by Fechner's Law.

Fechner's Law also makes an appearance. Chapter 63 applies Fechner's Law to loudness; here we apply it to frequency. It would predict that resonant frequencies are positioned logarithmically along the basilar membrane, and this is indeed a reasonable model. The top scale in Figure 71.2 shows the actual placement of the frequency resonances.[25] The scale is certainly far from linear: roughly the same length represents a 170 Hz interval near the apex and a 10000 Hz interval near the base. The bottom scale in the figure shows a true logarithmic scale. It is not a perfect match to the true scale, but it is an adequate model for the purposes of this book. Roughly, each centimeter along the cochlea corresponds to a decade, or multiple of ten, of frequency. Because of this convenient relationship, this book will use 3 cm as the basilar membrane length, even though that is on the short side of what is observed.

71c. Extra: Cochlear Details

Figure 70.1(a) shows some more detail of the inside of the cochlea. The line labeled (iv) represents the **vestibular membrane** (or **Reissner's membrane**), which means that the cochlea is actually divided lengthwise into three chambers: (i) the **scala tympani,** (ii) the **scala vestibuli,** and (iii) the **scala media** or **cochlear duct.** The liquid in the scala tympani and scala vestibuli is called **perilymph,** and it is these two chambers that are connected by the helicotrema. The cochlear duct is filled with a different liquid, the **endolymph.**

The hair cells are part of the **organ of Corti,** which sits on the basilar membrane on the cochlear duct side, surrounded by the endolymph. The difference between the endolymph and the perilymph on the other side of the basilar membrane is necessary for the hair cells to send out nerve impulses. The primary function of the vestibular membrane is to keep the two liquids separate; it is too thin to impact how the basilar membrane vibrates.

As sound travels along the cochlea from the oval window, it is a traveling wave in the liquid. Where it causes the basilar membrane to vibrate, that is also in the form of a traveling wave. (See Chapter 112 for what constitutes a traveling wave.) The description of the basilar membrane as vibrating in response to the wave is accurate, but it does not reveal this detail.

Where the sound wave reaches the resonating section of the basilar membrane, the membrane's vibration amplitude is large, but only relative to the amplitudes it caused on the rest of the membrane. Even large amplitudes are small fractions of a millimeter. Nevertheless, this is enough to transmit most of the sound energy from the scala vestibuli and scala media through the basilar membrane into the scala tympani. As a result, very little sound of that frequency continues towards the cochlear apex.

---

[25] Donald D. Greenwood, "A Cochlear Frequency-Position Function for Several Species—29 Years Later." *J. Acoust. Soc. Am.* 87(6) (1990): 2594.

Once the sound is in the scala tympani, most of the energy returns towards the base, to the round window. The flexibility of the round window allows that energy to be dissipated, so that the sound does not continue to travel around in the cochlea.

Chapter 72. Combination Tones

It is not a coincidence that the maximum frequency for hearing beats, around 15 Hz, is only slightly below the lowest perceivable pitch of 20 Hz. Chapter 41 notes that the superposition of two pure tones of equal amplitude is mathematically equivalent to a single pure tone with an amplitude that varies at the beat frequency. If the two original tones get further apart in frequency than 15 Hz, then the loudness variation becomes too rapid to perceive directly. However, the perception that something odd is going on does not vanish. The loudness variation transitions into a buzzing quality, and then as the frequency difference becomes larger than roughly 20 Hz, there can be a separate perceptible tone called a **difference tone,**

$$f_{\mathrm{diff}} = |f_1 - f_2| \quad . \tag{72.1}$$

This is also known as a **Tartini tone,** named after composer, violinist, and early music theorist Giuseppe Tartini, who reported the effect in the mid-1700s.[26] Helmholtz discovered that difference tones are the most easily perceived examples of **combination tones,** which are sounds that are heard with frequencies that are simple combinations of the original frequencies.[27] Other examples are the summation tone at $f_1 + f_2$, the second difference tone $2f_1 - f_2$, and so forth. All of these other examples are very difficult to hear.

For a graphical illustration of Eq. 72.1, look ahead to Figure 83.1. The lower part of that figure illustrates how the beats described in Chapter 41 line up with the difference tone described here. Beats are shown as a dashed line to indicate that they aren't heard as a pitch, but a difference tone does sound similar to a real low-frequency pitch.

Helmholtz argued that these combination tones arise from non-linearities in combining the sounds, which is a fancy way to say that superposition breaks. This would mean that even though Eq. 72.1 is the same as Eq. 41.3 for beats, the physical mechanism is very different. Superposition (Chapter 37), which gave us the simple combination of spectra in Chapter 46, relied on the model called linear response. Non-linearity simply means that model is insufficient. One of the possible consequences is that when two sinusoids combine, the combination spectrum gets additional peaks, for instance at the difference frequency. It is most likely to occur for very large amplitude vibrations, similar to the way that Hooke's Law is likely to fail for large deformations of a spring. Indeed, the difference tone is easiest to hear when the original tones are very loud. Non-linearities also seem likely in the complex structure of the ear, although Helmholtz and others have shown that they can occur when sounds combine outside the ear as well.

However, we now know that this is not the whole story. When the two frequencies are each sounded in different ears of one person, for instance through headphones, it is still possible to hear the difference tone. These **binaural difference tones** show that neural processing is also involved.

Chapter 73. Periodicity Theory of Pitch

Although the place theory of hearing is an excellent model, explaining the biological basis of hearing and many of its consequences, there are some auditory effects for which the place theory cannot account. One is difference tones, since they amount to hearing something that is not present in the sound spectrum. Another such effect is mentioned in Chapter 42: Ohm's Acoustic Law is not completely true, as in some cases changes in the initial phase of the partials in a complex sound can be perceived.

---

[26] Giuseppe Tartini, *Trattato di musica secondo la vera scienza dell'armonia* (1754).
[27] Helmholtz, *On the Sensation of Tone*, 152.

The nerve signal that is triggered by each hair cell, when it vibrates, consists of a series of pulses that travel along the nerves. These pulses are synchronized with the phase of the hair cell vibration. Thus, there exists a possible biological mechanism for the brain to receive information about the phase of the sound, and more generally about how the motion associated with each partial is varying in time. Perceptions of pitch which seem to rely on this information fall under the **temporal theory of hearing.** This is not exactly a competing theory to the place theory, which is undoubtedly correct. But it does add another dimension to the model for hearing. This is also called the **periodicity theory of pitch,** because it allows for the perception of a pitch based on identifying the period of a repetition in the sound stimulus.

Chapter 72 shows that it is sometimes difficult to distinguish through experiments whether a perception originates in this periodicity, or results from a non-linearity in the ear response. This book simply notes that there are some elements of perception that are more easily understood from the perspective of oscillations in time than from the perspective of a spectrum. So, they will be described that way, without worrying too much about properly ascribing their physiological origin to the temporal theory or the place theory.

Chapter 74. Virtual Pitch

Section 44a introduces the idea that a spectrum could be harmonic, but with a no partial at the lowest harmonic frequency, the so-called **missing fundamental.** When listening to a sound with such a spectrum, the perceived pitch matches the frequency of that missing fundamental. This is the phenomenon of **virtual pitch,** the perception of a pitch for which there is no partial at the corresponding frequency.

There are several possible causes for virtual pitch. One is that the virtual pitch is the difference tone of the other partials that are present in the sound. However, virtual pitch is usually much clearer to the listener than the typical difference tone. As another possibility, it has been suggested that the listener's mind is subconsciously searching for patterns in the sound spectrum. Thus, like someone inspecting a spectrum visually, pitch perception could be based on the regular pattern of the spectrum, regardless of the presence of individual partials. It is difficult to test the validity of this viewpoint, based as it is on subconscious thought.

But another possible explanation for virtual pitch arises directly from the periodicity theory. Figure 74.1 illustrates the superposition of two pure tones with frequencies related by $f_3 = \frac{3}{2} f_2$. That is, they are the second and third harmonics of the frequency $f_1 = \frac{1}{2} f_2 = \frac{1}{3} f_3$. As can be seen, the combined vibration has a period corresponding to that "fundamental" frequency, even though there is no partial at that frequency. Although Figure 74.1 shows only one specific example, the same result would occur regardless of the amplitudes or initial phases of the partials. More generally, given a spectrum with a missing fundamental, the period of the combined, complex vibration will be that of the missing fundamental frequency.

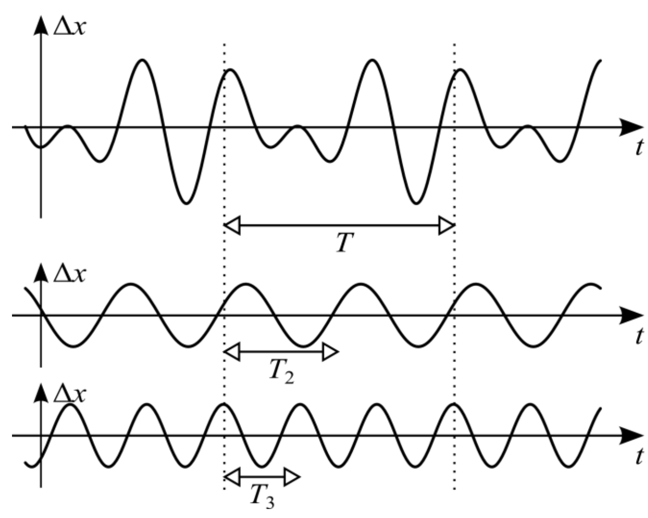

Figure 74.1
Superposition of harmonics without a fundamental.

Thus, the periodicity theory of pitch provides a direct explanation for the virtual pitch phenomenon. Regardless of what frequencies are present in the spectrum, the pitch may be perceived based directly on the period of the complex vibration.

Chapter 75. Repetition Pitch

Chapter 74 shows a case where a pitch can be heard without being present in the sound spectrum. But it is possible to hear a pitch in sound that is not even remotely periodic! The first careful observation of this phenomenon seems to come from Christiaan Huygens, who noticed the phenomenon in the sound of a fountain in the gardens of Chantilly castle in France.[28] The sound itself, directly from the fountain, was random noise. And yet, when combined with echoes of itself, each delayed in time due to having traveled further, the combination had an audible pitch.

The phenomenon, called **repetition pitch,** is most easily understood as being completely unrelated to sound spectra. Consider Figure 75.1, in which Δx_1 shows a random, nonperiodic motion representing some noise reaching your ear. The graph of Δx_2 is a duplicate of Δx_1, except that it is arriving at your ear Δt later (perhaps because of having traveled along a longer path from the same source). Vibration Δx_3 is again the same, delayed even later.

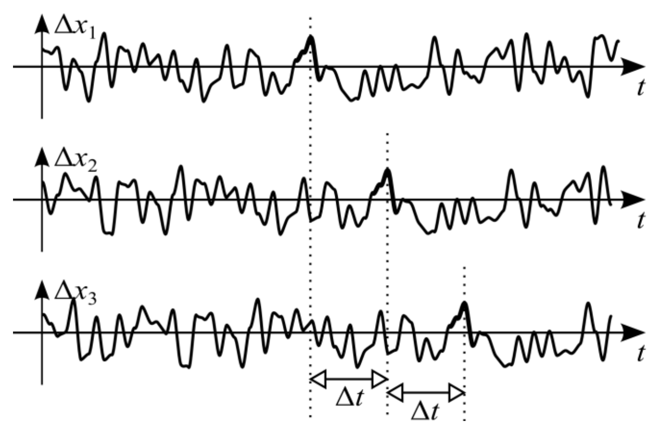

Figure 75.1

Random, unstructured motion, duplicated and shifted to later times.

The vibration actually experienced by your eardrum would be the superposition of these sounds. The superposition is not shown; you can imagine that finding the superposition would be a mess. However, certain prominent features might survive to be recognizable in the superposition; in Figure 75.1 one particular peak has been highlighted. Our hearing system can pick out those repeating prominent features and measure the time delay Δt between them. This time delay is then interpreted as if it were a period $T = \Delta t$, and the pitch for the corresponding frequency will be heard,

$$f = \frac{1}{\Delta t} \quad . \tag{75.1}$$

To be clear, the combined vibration never repeats, so there really isn't any period. But the pitch perceived is the same as if the time delay were a period.

This repetition pitch can be heard even if there is only one time-delayed copy of the original sound (such as only Δx_1 and Δx_2 from Figure 75.1). Not surprisingly, the effect is stronger if there are multiple, equally delayed copies. This was the situation that Huygens observed, because the sound from the fountain was echoing off a flight of stairs. Each step was further from him than the prior one by a uniform step size, so each echo was later by a uniform delay.

[28] Christiaan Huygens, "En Envoiant le Probleme d'Alhafen en France à." *Oeuvres Complètes de Christiaan Huygens* 10 (La Haye, Martinus Nijhoff: 1905): 570-571.

Chapter 76. Musical Intervals

Perceived pitch is almost entirely determined by frequency, essentially independent of sound intensity. If you have read Chapter 66, you know that the reverse is not true: perceived loudness depends on both intensity and frequency. But for many purposes, it is much more important to compare two pitches to each other, rather than to know the numerical frequency for a single pitch. A relationship between two pitches is called a **musical interval.**

Fechner's Law applies quite successfully to both the loudness and the pitch of sounds. For loudness, this leads to the invention of new scales with new units, described in Chapters 57 and 66. For pitch, similar scales have been proposed, but the most common way to work with pitches is less numerical. Indeed, the term "musical scale" refers to an ordered sequence of pitches, a meaning related to the mathematical "scale," but still quite different. The particular pitches chosen for a musical scale are called **notes.** The notes are assigned labels, such as *C*, *D*, *E*… (letter notation) or *do*, *re*, *mi*… (solfège system) or *Sa*, *Ri*, *Ga*… (Indian swara). But often the relationships between the pitches, i.e., the musical intervals, are more important than the pitches themselves. In fact, for swara, and sometimes for solfège, this is explicitly the case: *Ri* does not refer to any specific pitch, but instead refers to the note that is one step above whatever *Sa* is.

When comparing two pitches, there are certain pairs that are so similar that in every common musical scale the two members of the pair are given the same name. For instance, there are middle *C*, low *C*, and high *C*. The musical interval between two neighboring pitches in such a set is an **octave.** In line with Fechner's Law, increasing a pitch by one octave corresponds to multiplying the frequency, in this case by two. To describe this mathematically, we define the octave frequency ratio

$$R_o = \frac{f_B}{f_A} = 2 \tag{76.1}$$

where frequency f_B is one octave above frequency f_A, and the subscript is *o* for octave, not a zero.

In order to relate frequency relationships to your past experience, another handy musical interval is the **semitone.** Also called a **half step,** this is the smallest interval used in common music of Western culture. It is the interval between any two neighboring notes on a piano keyboard (including the black keys), or the interval between any two frets on a guitar. The frequency ratio for this interval can be determined by combining the precept of Fechner's Law with the fact that twelve equal semitone steps make one octave. So, starting at the same frequency f_A as above and multiplying by the semitone ratio R_s for each step up gives the equation

$$\left(\left(\left((f_A R_s)R_s\right)R_s\right)\cdots R_s\right) = f_A R_s^{12} = f_B = R_o f_A \quad , \tag{76.2}$$

$$R_s = \sqrt[12]{R_o} = \sqrt[12]{2} = 1.05946\ldots \quad . \tag{76.3}$$

A **whole tone** (or **whole step**) is a musical interval of two semitones, giving it a frequency ratio of

$$R_w = R_s^2 = 1.12246\ldots \quad . \tag{76.4}$$

Notice that combining musical intervals works through multiplication, and that changing frequency by several of the same interval implies raising the ratio to a power. Addition is never involved! This is probably different from the way you usually think about compounding numbers. And it is especially important because musicians often use words reminiscent of addition. To "increase by a semitone" would mean to multiply the frequency by R_s.

Musical intervals are best described as ratios between two frequencies. However, on occasion one might need to know the frequency difference instead. If f_h is one semitone higher than f_l, then their difference works out to be

$$\Delta f = f_h - f_l = R_s f_l - f_l = (R_s - 1) f_l \quad . \tag{76.5}$$

The frequency difference is proportional to the lower frequency. A similar calculation will show that the frequency difference is also proportional to the higher frequency, although with a different multiplier. The same result is found for any musical interval, simply by substituting in the appropriate frequency ratio.

Chapter 77. Pink Noise

On a logarithmic frequency axis, any given type of musical interval (octave, semitone, etc.) has a constant size, independent of what the lower frequency is. This shows up in the figures of Chapters 64–67, where octaves can be seen between 50 Hz and 100 Hz, 100 Hz and 200 Hz, and any other similar pairs. Contrast this with a linear frequency axis, on which musical intervals grow larger as the starting frequency grows larger.

Section 44c mentioned pink noise, which is more precisely defined as noise which has equal intensity in equivalent musical intervals. The idea of "intensity contained within a frequency range" will not find a lot of use in this book, since almost all of the spectra will consist of narrow peaks. Without getting into the mathematical details, you can consider the "intensity in a frequency range" to be proportional to the area on a spectrum that is both below the graph line and between the specified frequencies.[29]

Figure 77.1 shows two spectra, both of pink noise. In part (a), without a scale on the logarithmic frequency axis it is not possible to tell what musical intervals have been highlighted. But you can tell that they are the same, because they have the same width. Given that they have equal widths, they clearly have the same area. In part (b), even without a scale on the linear frequency axis, it is clear that both highlighted frequency ranges are octaves, because the high frequency of the range is double the low frequency. It is less obvious that the highlighted areas are equal, so that the graph satisfies the criterion for pink noise, but they are.

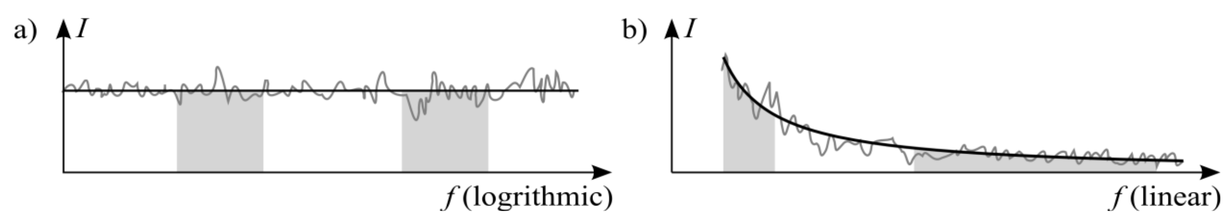

Figure 77.1

Pink noise spectra, on logarithmic and linear frequency axes.

Chapter 78. Steps Within a Range

78a. Frequency Increments

A range of frequencies, from a lower frequency f_l to a higher frequency f_h, is often called a **band** if it has some unifying feature and if the range is somewhat narrow compared to the frequencies involved. A band can be described by specifying its low and high limits, or by specifying its **center frequency** f_c and its **bandwidth** Δf. Those are related by the equations

$$f_c = \frac{f_l + f_h}{2} \quad , \tag{78.1}$$

[29] For the purist: In spectra without peaks, the vertical axis is not truly intensity, but **spectral intensity**. Describing spectral intensity properly is beyond the scope of this book. Suffice to say that in converting Figure 77.1(a) to (b), the lower frequency range gets compressed, resulting in the spectral intensity "bunching up," while the higher frequency range gets stretched and spectral intensity gets spread out.

$$\Delta f = f_h - f_l \quad . \tag{78.2}$$

For example, AM radio broadcasts are limited to the AM band, which runs from $f_l = 540$ kHz to $f_h = 1610$ kHz. That would be considered a rather wide band, since $\Delta f \approx f_c$. The FM radio band is between $f_l = 87.7$ MHz and $f_h = 108.0$ MHz. In this case $\Delta f \approx 0.2 \times f_c$, making it a band of moderate width.

Suppose that you have a set of frequencies in a band which are separated by a uniform increment Δf_s, which is smaller than the bandwidth. (The subscript *s* represents that this is a small *step* in frequency.) For example, the allowed frequencies for radio stations have a specified spacing. How would you determine the number of steps, or the number of frequencies, that can fit in the band? Since each step up the band adds to the frequency, if n steps are possible, then we can write

$$f_l + n\,\Delta f_s \lesssim f_h \quad , \tag{78.3}$$

where there is an inequality because the steps might not come out just right.

For example, suppose the band is from $f_l = 100$ kHz to $f_h = 255$ kHz, and the increment is $\Delta f_s = 8$ kHz. Solving this for n gives a result that you might have thought of from the beginning: dividing the range width by the unit interval,

$$n \approx \frac{f_h - f_l}{\Delta f_s} = 19.375 \quad \rightarrow \quad n = 19 \quad . \tag{78.4}$$

You can take 19 steps up from the lowest frequency. The answer was rounded down because the next step would go past the high frequency limit of the band. Note that with n steps, there is room for $n + 1 = 20$ different frequencies, because there is a frequency at both the beginning and end of the sequence of steps.

78b. Frequency Intervals

Suppose that you have a musical instrument that can produce a range of frequencies from $f_l = 100$ Hz to $f_h = 255$ Hz, which could be called the frequency band of the instrument. How many pitches on the Western musical scale could that instrument play? Here the step is a semitone, but musical intervals imply a frequency ratio, ***not*** a frequency difference. As with taking semitone steps up to an octave in Eq. 76.2, each step involves a multiplication instead of an addition. If the instrument allows for n steps, then the equation for multiple steps is

$$f_l R_s^n \lesssim f_h \quad . \tag{78.5}$$

Algebraically solving this equation for n requires logarithms. But if that seems difficult, a perfectly viable alternative is the "guess and check" method. Make a reasonable guess (for instance, maybe $n = 10$), and check the result,

$$f_l R_s^{10} = 178.18\ldots\ \text{Hz} \ll f_h \quad . \tag{78.6}$$

This guess was too small, so we make a larger guess, and then keep adjusting, homing in on the answer. A bad strategy would be to adjust the guess by only one; instead, try larger changes, to see how much they impact the result. With a good strategy, it is possible to find the answer quite quickly. In this example, $n = 16$ semitone steps can be made before exceeding the highest frequency. As in the previous section, this allows for $n + 1 = 17$ pitches on the instrument.

It is possible to ask questions for which "guess and check" is the *only* method available. For instance, if two people start at f_l and step up in frequency, one taking semitone steps and the other taking uniform Δf_s steps, after how many steps n would they both be at the same ending frequency f_e? The equation expressing this question is

$$f_l R_s^n = f_e = f_l + n\,\Delta f_s \quad . \tag{78.7}$$

This is called a **transcendental equation.** It literally impossible (not just difficult) to algebraically solve it for n! And there are other ways in which the practicing physicist can run into unsolvable mathematics. Sometimes physicists can find sneaky ways to get around the difficulty, by slightly changing the original question, or giving the answer a new name. (The ancient Greeks did something like this: when they couldn't solve for the ratio of a circle's circumference to its diameter, they gave it the name π.) But in the end, guess and check is a perfectly respectable method.

This section and the last section ask very similar questions, but one case requires the step increment and the other requires the step ratio. Sometimes, in real life as well as in academics, one of those is known but the other is more appropriate. Such situations are intrinsically imperfect, but they happen. To convert one into the other, consider the band's center frequency and a frequency f_{c+} one step (which might not be a semitone) higher. The increment and ratio can then be related by the equations

$$f_c + \Delta f_s = f_{c+} = f_c R_s \quad , \tag{78.8}$$

$$1 + \frac{\Delta f_s}{f_c} = R_s \quad , \tag{78.9}$$

$$\Delta f_s = f_c (R_s - 1) \tag{78.10}$$

Chapter 79. Intervals with Logarithms

If you wish to solve Eq. 78.5 directly with logarithms, the solution goes as follows:

$$\log(f_l R_s^n) \lesssim \log(f_h) \quad , \tag{79.1}$$

$$\log(f_l) + n \log(R_s) \lesssim \log(f_h) \quad , \tag{79.2}$$

$$n \lesssim \frac{\log(f_h) - \log(f_l)}{\log(R_s)} = \frac{\log\left({}^{f_h}/_{f_l}\right)}{\log(R_s)} \quad . \tag{79.3}$$

This kind of algebra with logarithms will not be used anywhere else in this book, but if you feel comfortable with it, it is more direct than the guess-and-check method.

Chapter 80. Critical Band

Each position along the basilar membrane has a single resonant frequency, as indicated in Figure 71.2. But each position responds with a significant vibration over a range of frequencies, for at least two reasons. Consider, as an example, an incoming pure tone of frequency 1.1 kHz and the reaction of the basilar membrane position that is tuned to 1.0 kHz. The first reason to expect that position to vibrate is that the basilar membrane is, after all, a membrane. Neighboring positions are connected by tissue. While the 1.1 kHz pure tone is causing the 1.1 kHz position on the membrane vibrate, some of the nearby positions have to vibrate as well, if the membrane is to avoid tearing. The second reason to expect the 1.0 kHz position to vibrate is that Figure 68.1 shows that when the driving frequency is only "close" to a resonant frequency, there can still be a responding vibration. The 1.1 kHz pure tone can still elicit a vibration of a vibrator with a 1.0 kHz resonant frequency. The response would not be as large as it would be for a 1.0 kHz tone, but it could still be significant.

The response to pure tone stimuli is illustrated in Figure 80.1. (The figure is using the idealized model with frequency laid out strictly logarithmically along the basilar membrane.) In response to a pure tone, a small region of the basilar membrane will vibrate with a variety of amplitudes. To simplify, we just say that a region, indicated by a pair of dashed lines, vibrates significantly. The region is roughly 1.2 mm wide, although the two given examples show that this varies quite a bit. This vibrating region is roughly centered on the position that resonates best with the pure tone stimulus.

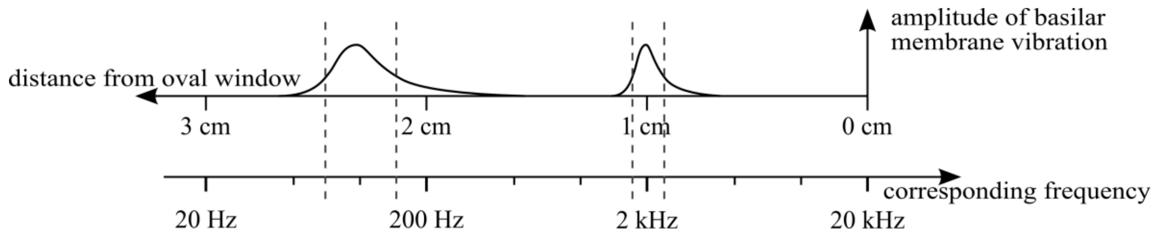

Figure 80.1

Illustration of critical bands at 100 Hz and 2 kHz, viewed as measuring the size of excited region along basilar membrane.

The positions on the edges of the strongly vibrating region have their own resonant frequencies. Those edge frequencies define the limits of a **critical band** for that center frequency. For example, the left bump in Figure 80.1 shows that the the point on the membrane that is 2.3 cm from the oval window resonates with 100 Hz. The width of that bump indicates that the critical band for 100 Hz runs roughly from 65 Hz to 150 Hz.

To take an entirely different perspective, we can consider the response of a single point on the membrane to a variety of stimulus frequencies. The result would be a graph somewhat like Figure 68.1. A full treatment from this perspective would require many such graphs, showing the response versus frequency for every position on the basilar membrane. But again, we can simplify by describing each position on the basilar membrane as responding to a range of frequencies. For example, the point on the membrane that is 2.3 cm from the oval window (with a resonance frequency of 100 Hz) will vibrate for any stimulus frequency between 65 Hz and 150 Hz. So this alternative view obtains the same critical band.

Each position on the basilar membrane has its own critical band, with a center frequency and a bandwidth. Those two parameters vary smoothly from one end to the other. The critical band idea therefore can be described by specifying both the center frequency and bandwidth over the whole length of the basilar membrane. Figure 80.2 plots measured critical bandwidths as a function of the central frequency (the black curved line).[30] The two ends of the curve correspond to the physical ends of the basilar membrane, but positions are not directly represented in the graph. Notice that both axes have logarithmic scales, with the result that the musical intervals semitone and octave, included for comparison, form straight lines.

Although the critical bandwidth Δf_{cb} varies smoothly with center frequency, it is fairly constant at the lower center frequencies, and then for higher pitches it rises roughly parallel to the musical intervals. The gray straight line indicates a rough model for this, described by the equations

$$\Delta f_{cb} = 88\ \text{Hz} \quad (\text{for } f_C \leq 434\ \text{Hz}) \quad \text{and} \tag{80.1}$$

$$\Delta f_{cb} = 0.20\, f_C \quad (\text{for } f_C \geq 434\ \text{Hz}) \quad . \tag{80.2}$$

Keep in mind that these equations do not describe the actual critical bandwidth function. But they do form the best model that can be made from simple pieces. The two relationships have implications for the question of counting how many critical bands are in a larger range, such as an octave. Below 434 Hz, the constant critical bandwidth (Eq. 80.1) calls for forms like Eq. 78.3, while above 434 Hz, because the critical bandwidth is proportional to the center frequency (Eq. 80.2), forms like Eq. 78.5 are needed.

Critical bands are important for understanding what is heard in various situations involving two sounds. For example, when two sounds are close enough for their critical bands to overlap, the combination results in a rough, slightly unpleasant perception. Another example is that in the presence of a loud sound, a second softer sound within its critical band might not be heard at all, a situation called **masking.**

[30] Hartmut Traunmüller, "Analytical expressions for the tonotopic sensory scale." *J. Acoust. Soc. Am.* 88 (1990):99–100.

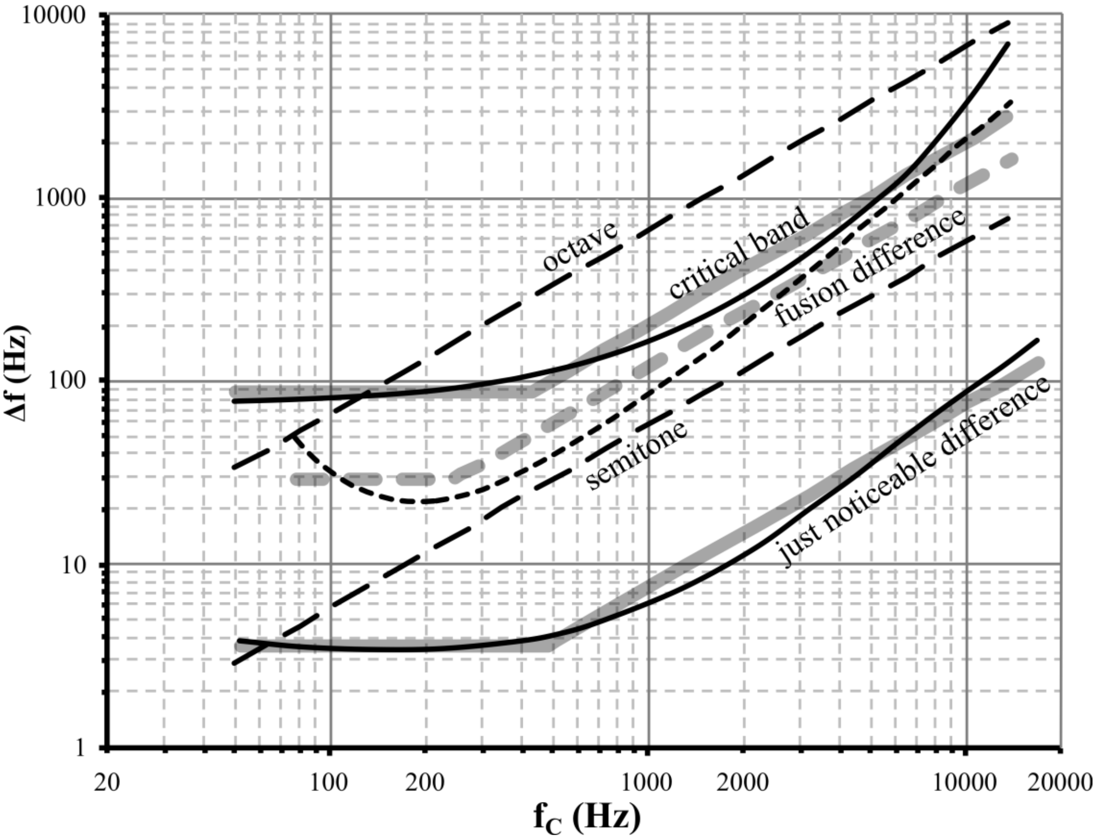

Figure 80.2

Critical bandwidth (using equations from Traunmüller[30]), fusion difference (based on data from Plomp[32]), and just noticeable difference (based on data from Fastl and Zwicker[31]) as a function of band center frequency. Thick gray lines are models for respective curves consisting of either constant Δf or constant $\Delta f / f$. Semitone and octave shown for comparison.

There are a number of details which this book will not get into. For instance, the responses that define the critical bands are not really symmetric about their peak, either in basilar membrane position (as you can see in Figure 80.1) or in frequency. Another is that there is an active process in the living ear that narrows peaks such as shown in Figure 80.1. Since some experiments on this subject are performed with ears from cadavers, where this process is not active, published critical bandwidths can vary quite a bit. The critical bandwidths in Figure 80.2 are determined from experiments based on perception by live subjects, and thus avoid this complication.

Chapter 81. Frequency Just Noticeable Difference

If two sounds are the same except for a very small difference in frequency, then they may be perceived as being the same. This chapter considers the situation where such sounds are presented sequentially, and Chapter 83 looks at two simultaneous sounds.

Suppose two pure tones, f_l and f_h, when sounded one after the other are just barely perceptible as having different pitches. The difference between their frequencies is the **frequency just noticeable difference,** also called the **frequency difference limen.** This book will abbreviate that as the **fjnd**, to distinguish it from the intensity jnd. Figure 80.2 plots the fjnd as a function of the center frequency f_c (halfway between the two tones).[31] The fjnd is quite small. At most frequencies it is much smaller than a semitone, which is included in the figure for comparison.

Although the fjnd varies smoothly with center frequency, it is fairly constant at lower frequencies, and then rises for higher pitches roughly parallel to the musical intervals. The light gray line indicates a rough model for this, described by the equations

$$\Delta f_{jnd} = 3.6\text{ Hz} \quad (\text{for } f_C \leq 487\text{ Hz}). \tag{81.1}$$

$$\Delta f_{jnd} = 0.0074\, f_C \quad (\text{for } f_C \geq 487\text{ Hz}). \tag{81.2}$$

Keep in mind that these equations do not describe the actual fjnd, but they form the best model that can be made with the two simplest types of relationships.

Notice the implications of these relationships for counting how many just noticeable differences are in a larger range. Below 487 Hz, the constant fjnd calls for forms like Eq. 78.3. Above 487 Hz, since the fjnd is proportional to the frequency, it should be treated as a musical interval with forms like Eq. 78.5. We can derive the appropriate ratio by combining Eq. 81.2 and Eq. 78.10.

$$\Delta f_{jnd} = 0.0074\, f_C = 0.0074\left(\frac{f_l + f_h}{2}\right) \tag{81.3}$$

$$f_h - 0.0037\, f_h = f_l + 0.0037\, f_l \tag{81.4}$$

$$R_{jnd} = 1 + \frac{\Delta f_{jnd}}{f_c} = 1 + 0.0074 = 1.0074 \quad . \tag{81.5}$$

That ratio then sets us up to find the number of just noticeable difference in a range. For example, to find how many just noticeably different pitches are in the octave 1 kHz to 2 kHz, the appropriate equation is

$$2\text{ kHz} = R_{jnd}^{n}(1\text{ kHz}) \quad . \tag{81.6}$$

But keep in mind that form is appropriate only because the frequencies involved are higher than 487 Hz.

Chapter 82. Fjnd and Critical Bands

One very noticeable feature of Figure 80.2 is that the jnd curve has a shape that is extremely similar to the critical bandwidth curve. We can seek to understand the reason by considering how the basilar membrane vibrates in response to a pure tone. Figure 82.1 illustrates the response to two very similar, but distinguishable, frequencies.

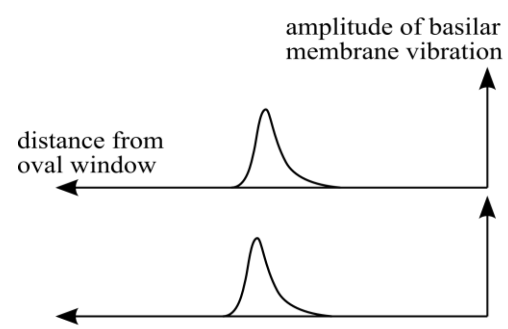

Figure 82.1

Vibrations on basilar membrane due to very similar frequencies occurring at different times.

An fjnd is much smaller than a critical band; there are roughly 22 jnds in one critical band, depending on center frequency. Some sources use this fact to argue that another frequency detection mechanism (which is covered in Chapter 73) must be involved. However, that would not explain the similarity between the fjnd and critical bandwidth curves.

In Figure 82.1, consider this: how close together could those two peaks get before you could not tell the difference between them visually? Clearly this depends on the shape, with narrower peaks being easier to distinguish. But equally clearly, the shift required to distinguish them is much smaller than the width of the peak. This example differs from hearing is a few ways. It is

[31] Hugo Fastl and Eberhard Zwicker, *Psychoacoustics: Facts and Models* (Berlin: Springer, 2007), 183.

based on sight, while hearing is more akin to touch (feeling something on different positions of the basilar membrane). Also, the two peaks in Figure 82.1 are separated by a vertical distance, while the two sounds in an fjnd experiment are separated by time.

Nevertheless, this thought experiment helps to clarify why the fjnd can be both closely related to, and yet so much smaller than, the critical bandwidth. Wider critical bands would require a larger shift in order to be readily distinguished. But the shift required is significantly smaller than the bandwidth itself.

Chapter 83. Fusion Frequency Difference

If two pure tones of similar frequency are heard simultaneously, the situation is quite different from sequential perception. If you have read Chapter 41, you have seen how two such sounds, if very close in frequency, can physically superpose to make a single sound with beats of the amplitude. But even if the two frequencies are too far apart for that result, they may still be perceived as a single pitch.

The separation required to barely distinguish two simultaneous pure tones is called the **auditory fusion frequency difference.** As long as the actual difference is less than this fusion frequency difference, it will sound as if the two sounds are fused together into a single sound with a perceived frequency that is the average of the two true frequencies,

$$f_c = \frac{1}{2}(f_1 + f_2) \quad , \tag{83.1}$$

The upper part of Figure 83.1 illustrates this. The fused sound may also have a rough or buzzing quality.

Figure 80.2 shows how this fusion frequency difference depends on the average of the two frequencies, as determined in experiments by Plomp.[32] For questions about how many fusion frequency differences are in a larger range, we only have two models to work with: a difference that is constant, and a difference that is proportional to the center frequency. The light gray line indicates a model for the curve, described by the values

$$\Delta f_{\text{fuse}} = 29\text{ Hz} \quad (\text{for } f_C \leq 245\text{ Hz}) \quad , \tag{83.2}$$

$$R_{\text{fuse}} = 1.12 \qquad (\text{for } f_C \geq 245\text{ Hz}). \tag{83.3}$$

This is clearly not a great model, but it is the best that can be made based on the available methods.

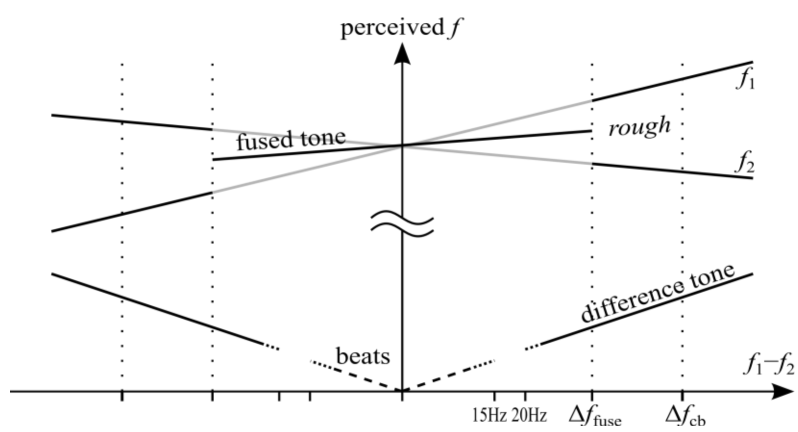

Figure 83.1

Various effects observed when two pure tones have similar frequencies, as a function of the difference between those frequencies. Individual frequencies f_1 and f_2 are arbitrarily chosen, as effects shown depend only on their difference.

[32] R. Plomp, "The Ear as a Frequency Analyzer." *J. Acoust. Soc. Am.* 36(9) (1964): 1628–1636.

There is another effect with a similar name, **auditory flutter fusion.** This involves a sound that is rapidly varying, for example being intermittent. If the variation is fast enough, then the perception fuses into a single steady perception. Despite the similar name, this is quite a different phenomenon from auditory fusion.

Chapter 84. Fusion Difference and Critical Bands

The fusion frequency difference changes with center frequency in a way somewhat similar to the critical bandwidth. The connection between them can be understood qualitatively. Figure 84.1 shows two situations with two pure tones exciting the basilar membrane simultaneously. As long as they are not extremely loud, the resulting membrane vibration will be given by superposition.

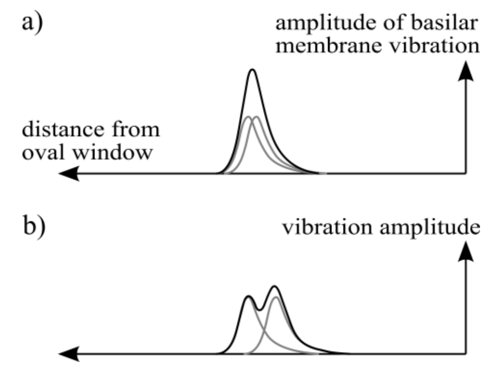

Figure 84.1

Vibration on basilar membrane due to two tones of very similar frequency occurring at the same time. Gray curves are responses to the tones separately.

In part (a) of the figure, the two tones are so close that their superposition forms a single peak. The ear has no information about the underlying reason for this, so only a single pitch is perceived. This can happen even if the pitches are more than a frequency just noticeable difference apart.

When the two frequencies are farther apart, as in part (b) of the figure, the superposition has two peaks. the auditory system has sufficient information to tell that the basilar membrane vibration is due to two components. The fusion frequency difference is the minimum difference required for this to occur, so that two separate pitches are heard. The shape of the basilar membrane excitation, and especially its width, is a critical determinant of that frequency difference is required.

Once the frequencies are further apart than the fusion frequency difference, the critical bands continue to overlap, until the two pitches are one critical band apart. If the frequency difference is in this range, there will be a roughness in the combined sound. This roughness diminishes as the frequencies get further apart, disappearing when the critical bands cease to overlap, as illustrated by Δf_{cb} in Figure 83.1.

Chapter 85. More Musical Intervals

85a. Fifth and Fourth

Chapter 76 introduces the idea of a musical interval, a relationship between two pitches. Because of Fechner's Law, a musical interval is best described mathematically as a ratio between two frequencies. But to understand the mathematical underpinnings of music, we need to define more musical intervals than the octave and semitone.

Certain pairs of musical pitches sound much nicer to humans, when played simultaneously, than other pairs. Such a pair is called a **consonance,** and the two pitches are said to be **consonant** with each other. The opposite, two pitches that sound grating when heard together, are said to be **dissonant.** According to legend, it was **Pythagoras**, a Greek philosopher of the 6th century BC, who discovered that consonant pitches are related to simple ratios. Pythagoras probably noticed these ratios in the lengths of strings on musical instruments. Chapter 179 explains why those lengths are closely related to the frequencies. In any case, the resulting rule is as follows.

> Two musical pitches are consonant if the musical interval between them, i.e., the ratio of their frequencies, is equal to a ratio of two small positive whole numbers. Smaller whole numbers result in more consonant pairs of pitches.

Thus, the most perfect consonance occurs between two sounds with the same frequency, called being **in unison**, so that

$$\frac{f_B}{f_A} = \frac{1}{1} \quad . \tag{85.1}$$

But that doesn't really count as a musical interval, since the two pitches are the same. The next best consonance is the octave, for which

$$\frac{f_B}{f_A} = R_o = \frac{2}{1} \quad . \tag{85.2}$$

Restricting ourselves to pitches that are closer together than an octave, the next best possible consonances are the **perfect fifth,**

$$\frac{f_B}{f_A} = R_{P5} = \frac{3}{2} \quad , \tag{85.3}$$

and the perfect fourth,

$$\frac{f_B}{f_A} = R_{P4} = \frac{4}{3} \quad . \tag{85.4}$$

Both of these musical intervals are often referred to as simply a **fifth** and a **fourth.** There are other musical intervals in use that are almost the same ratios and share those names, but they are **imperfect** because their ratios involve larger whole numbers, or may even be irrational (that is, they cannot be expressed as a ratio of whole numbers).

Notice that although the names are suggestive of the numbers 5 and 4, which have even been used as subscripts, those numbers have almost nothing to do with the frequency ratios. Those names, along with the less conspicuous "eight" associated with the word octave, derive from how many notes separate these pitches in a standard musical scale. But the numerical names are of no use whatsoever in remembering the ratios that they represent. It is best to simply think of fifth and fourth as arbitrary names.

As a reminder from Chapter 76, changing frequencies by an interval is achieved by multiplying or dividing by the appropriate ratio. Thus, compounding intervals means using several multiplications, even though in music theory the language of addition is common. For instance, increasing from frequency f_A by "a fifth plus a fourth" takes us to the new frequency

$$f_B = f_A R_{P5} R_{P4} = \left(f_A \frac{3}{2}\right)\frac{4}{3} = f_A \frac{4}{2} = 2f_A \quad , \tag{85.5}$$

by which we conclude that a fifth and a fourth "add together" to make an octave.

Table 85.1

Musical intervals of progressively less consonance.

Row	Musical Interval Name	Frequency Ratio
a)	just major sixth	$R_{M6} = \frac{5}{3}$
b)	just major third	$R_{M3} = \frac{5}{4}$
c)	just minor third	$R_{m3} = \frac{6}{5}$
d)	harmonic seventh	$R_{h7} = \frac{7}{4}$
e)	lesser septimal tritone	$R_{lst} = \frac{7}{5}$
f)	septimal minor third	$R_{sm3} = \frac{7}{6}$
g)	just minor sixth	$R_{m6} = \frac{8}{5}$
h)	septimal major second	$R_{sM2} = \frac{8}{7}$
i)	just major second	$R_{M2} = \frac{9}{8}$

85b. Extra: Less Consonant Intervals

Continuing to find small whole numbers with which to make ratios that are less than an octave, we find the musical intervals listed in Table 85.1. Working down the table, the intervals are less and less consonant, until somewhere between rows (d) and (g) we have crossed into dissonance. Thus, we find that in the rule for consonance, "small whole numbers" means roughly numbers less than six.

Although Table 85.1 is arranged in increasing order of the numbers involved, it should not be taken to be a strict progression from consonance to dissonance. Opinions about the consonance of even the first three intervals in Table 85.1 are subjective, and have varied throughout history and depending on culture. In music theory they are called **imperfect consonances**, as opposed to the "perfect" fifth and fourth. The intervals involving the number seven are rarely encountered in practical music.

Chapter 86. Dissonance

Why are certain musical intervals consonant? Actually, consonance is the absence of dissonance, so a better question is, what causes dissonance?

Studies have found that the consonance rule in Section 85a actually does not apply for pure tones. For pure tones, dissonance only occurs when the two tones are within one critical band of each other.[33] Notice that this makes it easier for lower pitches to be dissonant, because the critical bandwidth becomes proportionally larger for lower frequencies. The dissonance arises because the two regions on the basilar membrane that are being excited overlap. In combining, rapid beats occur, resulting in the dissonance sensation. If the two pure tones are exactly the same frequency, however, the dissonance disappears, because they combine by superposition into a single pure tone.

It is important that the rule for consonance in Section 85a referred specifically to *musical* pitches. From Section 44a, these are typically harmonic sounds, with overtones at frequencies that are multiples of a

[33] R. Plomp and W. J. M. Levelt, "Tonal Consonance and Critical Bandwidth," *J. Acoust. Soc. Am.* 38(4) (1965): 548–561.

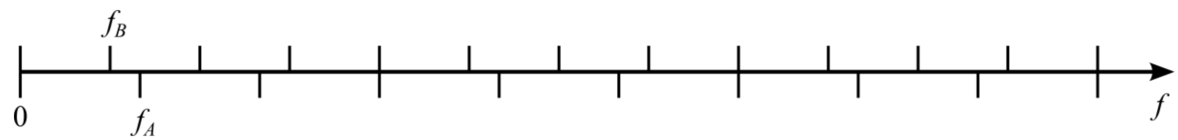

Figure 86.1
Harmonics of frequencies separated by a perfect fourth. Harmonics either match or are well separated.

fundamental. Even if the fundamentals of two musical pitches *are not* close in frequency, they may have overtones that *are* close in frequency. Their resulting excitations in the ear can overlap and cause dissonance. But if the overtones are extremely close, then the excitations can merge, avoiding dissonance.

Suppose that a musical sound with frequency f_B is one octave above a musical sound with frequency f_A. Then *every* partial of sound *B* will have exactly the same frequency as an overtone of sound *A*,

$$f_{Bn} = nf_B = n(2f_A) = f_{A(2n)} \quad . \tag{86.1}$$

Specifically, the n^{th} harmonic of *B* matches the $2n^{\text{th}}$ harmonic of *A*. With no combined partials that are nearly (but not exactly) the same, there is no dissonance. As a result, the octave is a very consonant interval.

If f_B/f_A is instead a ratio of other small numbers, then the frequencies of some, but not all, of their overtones will match exactly. Specifically, if the ratio is expressed as a reduced fraction, then the numbers in the fraction tell how often they match. For example, for the perfect fourth illustrated in Figure 86.1 $\left(\frac{f_B}{f_A} = \frac{3}{4}\right)$, one in four harmonics of B will match a partial of A, and one in three harmonics of A will match a partial of B. Thus, smaller numbers in the fraction result in more matches. Also, larger numbers in the fraction result in "near misses" which are closer together, creating opportunities for dissonance between those overtones. This is the mathematical and physiological origin of the rule for consonance.

These considerations have been based on the assumption that each musical sound includes all harmonics in its spectrum. This is an appropriate assumption for making a general statement about all musical sounds. But for any specific instrument, some or even many of the harmonics may not be present, as reflected in the instrument's timbre. An instrument that tends to only have a few harmonics in its spectrum is less likely to create dissonance than one with lots of harmonics in its spectrum.

Chapter 87. Musical Scales

Choosing which pitches to use in music and placing them in order by frequency defines a **musical scale.** In order to make pleasing music, choosing the pitches for a scale becomes an exercise in selecting the intervals between the pitches so that there are many opportunities for consonant intervals. As it happens, there is a near-coincidence between the two most consonant intervals that drives the choice of scales towards 12 intervals per octave.

Changing pitch by 12 fifths is nearly the same as changing by 7 octaves, as seen in the relationships

$$R_{P5}^{12} = \left(\frac{3}{2}\right)^{12} = 129.7463 \quad \approx \quad R_o^7 = 2^7 = 128 \quad . \tag{87.1}$$

For comparison, a piano spans just over 7 octaves. These relationships mean that if you started at the lowest note on a piano, and stepped up 12 perfect fifths (which would not be available on a normal piano, for reasons that we will soon see), you would arrive at a frequency, near the high end of the piano, which is less than one-quarter of a semitone different from the pitch 7 octaves above the that lowest note.

This is illustrated in Figure 87.1(a). This figure uses logarithmic axes, so that equal frequency multiples appear as equal spacing on the axis. Thanks to Fechner's Law, that is what "equal" means in this context. In part (a) the rightmost two hash marks, above and below the axis, don't quite line up. The sequence of intervals above the axis is known as the **circle of fifths.** It is a circle in the sense that it has returned (almost) to the same note that it started from—recall that pitches separated by octaves are so similar to each other, that they are given the same note name.

Had this math worked out perfectly, there would be no question about the best choice of frequencies for our scale. As shown in Figure 87.1(b), for each of the pitches in the circle of fifths we find the octave-equivalent pitch in the central octave. The gray arrows illustrate shifting six of the pitches by multiplying or dividing by powers of two, which on the log scale means shifting by multiples of the octave width. The result is an octave with twelve distinct notes, called a **chromatic scale**. The particular method in Figure 87.1(b) is called Pythagorean tuning. The upper hash marks have been given different heights, to help identify where they move to in the central octave. Hash marks of the same length are from the same initial octave, and therefore shift by the same amount.

Unfortunately, because the two ratios in Eq. 87.1 are not quite equal, none of this quite works. The note-to-note intervals within the central octave are not quite equal, which is barely visible in Figure 87.1(b). Some pairs of notes, which should be consonant, are instead dissonant. And so, from this mathematics of consonance arose many centuries of frustrated musicians, who could never quite get the perfect scale. The problem is not so severe for single-line melodies, but for music with chords and harmonies some combinations just don't sound good. Although some musical instruments could adjust pitch slightly on the fly, the advent of instruments with fixed pitches, especially the pipe organ and piano, meant that solution was no longer available.

Many adjustments to this basic scale have been tried throughout history in many cultures. Historical musical cultures in many areas of the globe include scales of a subset of those 12 pitches per octave. This reduces the problem, but it also limits the possible musical complexity. In Western culture, starting especially in the 1500s, different solutions were explored that fall in the category of **just intonation.** In

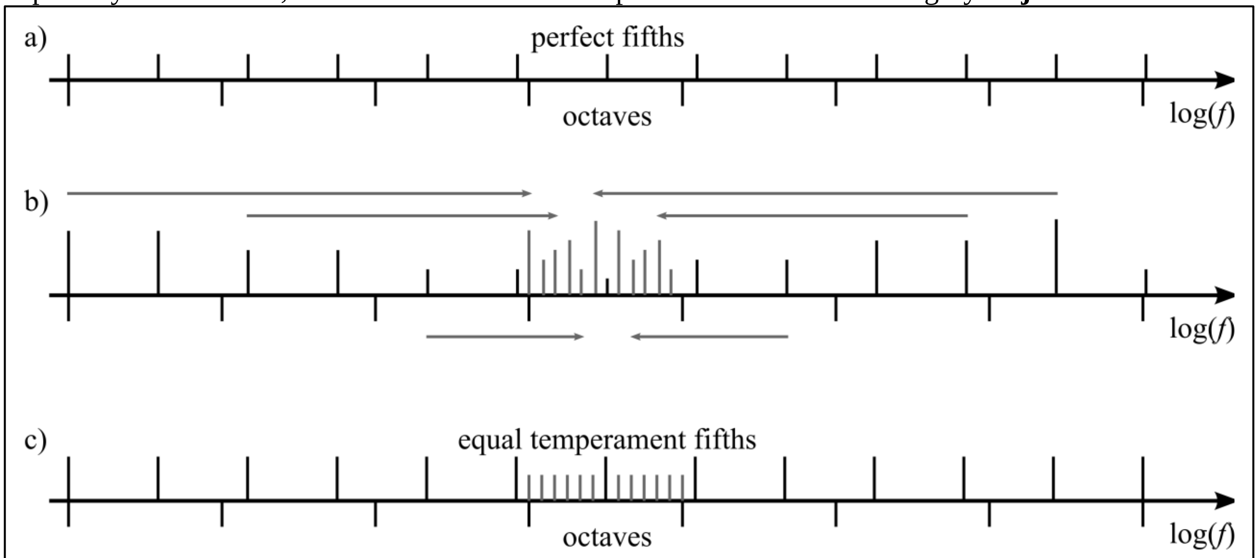

Figure 87.1
Relationship between octaves and fifths, both perfect and equal temperament.

these scales, some of the pitches were adjusted to frequency ratios with smaller whole numbers, so that they became more consonant. Inevitably, this means that other pairs of notes became more dissonant. The strategy was to compose music so that it avoided these dissonant intervals.

Another approach, explored over about the same time period in both Western and Chinese culture, was called **temperament.** Here intervals are adjusted away from small-number frequency ratios, but to a small enough degree that the change is not very noticeable. This has ultimately led to the system used nearly universally in Western music since roughly 1800, the **equal temperament** scale. As shown in Figure 87.1(c), all the fifths are equally tempered very slightly smaller so that the near miss at the end of the circle of fifths in Figure 87.1(a) is transformed into an exact hit. As shown in the central octave, when this circle of tempered fifths is gathered together, a scale results with all equal intervals between neighboring notes. This is the origin of the semitone, introduced in Chapter 76. With the equal temperament scale, none of the intervals available (except the octave) are perfect consonances. However, all the intervals that are expected to be consonant are very close to perfect.

When the pitches in a scale are not equally spaced, it is very important to identify the pitch at which each octave starts, called the **tonic**. Equal temperament has the advantage that since the notes are all equally spaced (on a logarithmic frequency axis), it makes no difference which one is considered the start of the octave. Figure 87.2(b) shows one way that letter names are associated with the pitches of equal temperament, but the choice to start with C is arbitrary.

On the Indian subcontinent, yet another solution is commonly used. Instead of trying to adjust the pitch of the twelve notes, 22 pitches are named in each octave. The majority of these come in pairs that are very close in pitch, nearly a just noticeable difference apart. In the naming system shown in Figure 87.2(a), the pairs use the same letter. The spacing of these notes is chosen so that for every interval that should be consonant, choosing the proper note from each pair can give perfect consonance (a frequency ratio with small whole numbers). A melody would never use both pitches from a pair, as if they were separate notes. But whenever harmony is required, notes are chosen from the pairs appropriately.

There are other scales used by different cultures, including a few that do not have a connection to the circle of fifths. But the consonance of the fifth, along with the remarkable mathematical coincidence of the circle of fifths, have resulted in a large majority of the world's cultures focusing in on slight variations of this twelve-tone scale. Outside of this chapter, all discussions in this book are in the context of the equal temperament scale.

Everything in this chapter so far has explored the musical intervals between notes, but has not chosen a specific frequency for any note. That choice is of far smaller consequence. In Western tradition, there have been several conventions. It is curious to note that over history, the frequency assigned to any specific note

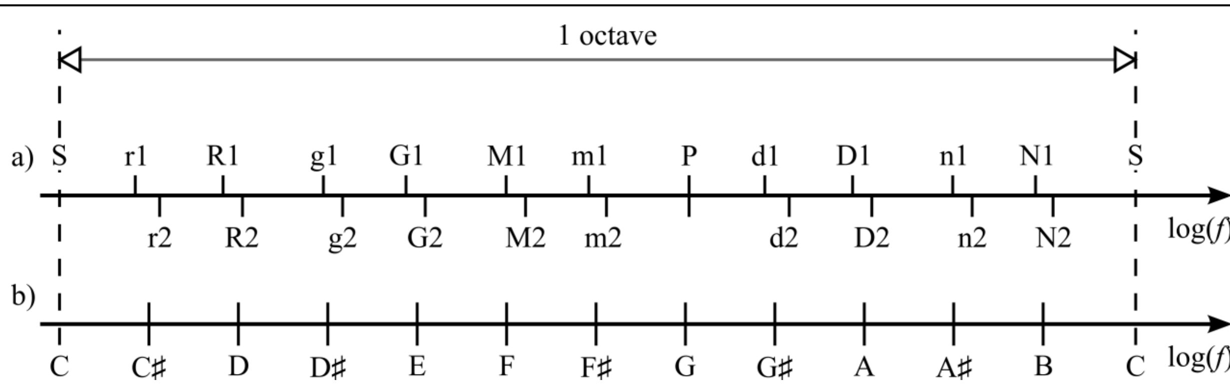

Figure 87.2

Ways to define intervals within an octave. (a) System used in music of the Indian subcontinent. (b) Equal temperament.

has slowly risen. The most common modern choice is to assign 440 Hz to the note A; the intervals of the scale then determine all the other frequencies. The Indian tradition removes the question by having no convention at all. The Indian scale is completely free-floating, with the frequency assigned to S, the tonic, varying by performance.