Part F: Shaping Sound

Chapter 88. Sound Signals

88a. Transducers and Amplifiers

In prior chapters, this book has relied on the close association between sound and vibration (displacement versus time) to investigate the time varying aspect of sound. Vibrations are often represented by other varying quantities instead of displacement, for instance, voltage versus time in a speaker cable, or frequency versus time in an FM radio wave. The physical quantities in those examples (voltage, EM field) are studied in the field of electromagnetism, but you don't need to know much about electromagnetism in order to think of them as representing the displacement in a vibration or sound. Any quantity that varies with time to represent a sound, this book will call a **signal**. The physical quantity that represents the displacement determines what **type of signal** it is.

When sound is recorded, the time dimension of sound also has to be represented by some other quantity. Often that other quantity is position. Examples of a **stored signal** might include magnetization versus position on an audiotape, or even displacement versus position in the graphs that are drawn throughout this book.

In either case, if the representing quantity is proportional to the represented vibration displacement, then it is a **proportional signal.** The alternative, a **non-proportional signal,** may be more difficult to work with mathematically, but it is not intrinsically any worse for representing the vibration. In fact, in some cases a non-proportional signal might be better; the Dolby noise reduction systems for magnetic audiotapes are good examples. But this book will only consider proportional signals.

Signals can be either **digital** or **analog.** In digital signals, time (or its representative in a stored signal) is divided into small but finite intervals. For instance, in the standard for Compact Discs, the time intervals are 22.73 μs long. Within each of these time intervals, the quantity representing displacement can only have one of a set of allowed discrete values, as shown in Figure 88.1. In this way, a digital signal *approximates* a vibration or sound; a digital signal *never* represents the vibration perfectly. The great benefit of digital signals is that they can be preserved perfectly as the signal moves from one place to another. During transmission, if some problem causes one of the values to deviate from that allowed discrete set by a small amount (that is, by less than one step size), then the error can be noticed and corrected.

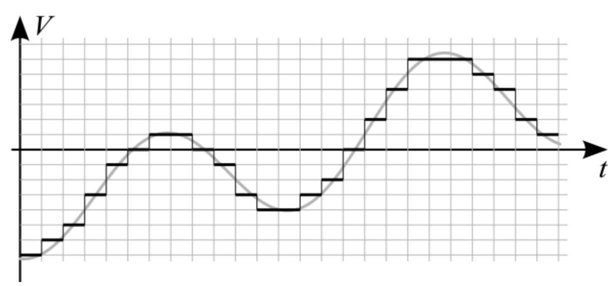

Figure 88.1

Digital voltage signal (black steps) and analog voltage signal (gray curve) representing a vibration. In each time interval, digital signal represents vibration by one of the horizontal grid lines.

An analog signal, on the other hand, allows the quantity representing displacement to have any value at all, and allows it to vary continuously with time (or its representative in a stored signal). An analog signal has the potential to represent a vibration or sound perfectly. On the other hand, if even the smallest error creeps into an analog signal, there is no way to tell (from the signal itself) that the error did not come from the original sound.

Sometime near the beginning of the 21st century digital signals became the preferred form in many situations, especially for stored signals. But real sounds, our sensory organs, and the real world are not digital, so analog signals will always play some role in processing sounds. This book will look at the basics of how analog signals can be affected as they carry representations of sound.

A **transducer** is anything that changes a signal of one type (the **input**) into a signal of another type (the **output**), while representing the same sound. For example, an eardrum is a transducer from a sound-in-air signal into a vibration signal, and a radio receiver is a transducer from a radio wave signal into an electrical voltage signal. On the other hand, any device that translates an input signal into an output signal of the *same* type, but normally with a different scale, is called an **amplifier.** Sometimes the word amplifier is reserved for a device that makes the signal larger, and a device that reduces the signal is called an **attenuator.** But that distinction can get awkward. A single device can be adjustable (with a volume control, for instance) to either increase or decrease the signal. This book will use the term **amplifier** for either case, in the same spirit that physicists use the term acceleration to describe a changing velocity regardless of whether it is increasing or decreasing.

Many ideas apply equally well to both transducers and amplifiers, so this book will use the generic term **device** to mean either of them. A device is described as **ideal** if the output signal faithfully represents the same vibration or sound as the input signal. For the purposes of this book, this "ideal" means mathematically that the output is proportional to the input; there are other options for faithful representation, but they are more mathematically complex. In the following Chapters 89–95 and 109–111, we'll have a look at how various transducers or amplifiers may deviate from being ideal.

One particularly important signal type to consider is the electrical signal. This book will not cover the physics involved with electricity to a great degree. The main electrical quantity needed will be **electrical potential,** measured with the root unit **volt** (abbreviation V). You are probably familiar with batteries that "have" 1.5 V. What this means is that one end of the battery is 1.5 V higher in electrical potential than the other end, which highlights one characteristic of electrical potential: only **electrical potential *differences*** are ever physically meaningful. This is why electrical connections most often have (at least) two wires, including power plugs and speaker cables. It is the electrical potential *difference* between the wires that represents the signal in such cables. Because "electrical potential difference" is such a mouthful, it also has the shorter name **voltage.**

Chapter 103 goes into more detail about electrical behavior, so that the operation of various types of microphones and speakers can be described. But you can do a great deal knowing only that voltage is an electrical concept that can represent a sound signal, and that microphones and speakers are transducers between voltage signals and sound.

88b. Response

For any transducer or amplifier, the **response** is the ratio of the output signal to the input signal. For an ideal device, the response has a single numerical value. When a device is not ideal, then the output-input ratio might vary, for example, with different input amplitudes or different frequencies. But to start with, we will consider the response of an ideal device.

Since a transducer changes the type of signal, the input and output are likely to be measured with different units, so that the response has units as well. For instance, suppose that the output of a microphone is a voltage oscillation that is proportional to the air pressure of the sound at the microphone input. (Air pressure is explained in Chapter 125, but here you can just consider it to be some physical quantity that oscillates to make sound.) In particular, suppose that a pressure amplitude of $0.2\,\frac{\mathrm{N}}{\mathrm{m}^2}$ causes a voltage amplitude of 5.0 mV. The response of that microphone is then

$$\text{response} = \frac{\text{output}}{\text{input}} = \frac{5\ \mathrm{mV}}{0.2\ \mathrm{N/m^2}} = 0.025\,\frac{\mathrm{V}}{\mathrm{N/m^2}} \quad . \tag{88.1}$$

You may not be familiar with the units in Eq. 88.1, but that isn't necessary here. The main point is that there are many different possible units for all the different kinds of transducers.

The previous example obtained the response from the amplitudes, or maximum values, of the input and output signals. This is a convenient point of comparison. But since the signals are proportional, the response is the ratio of the signal quantities at every point in time, not just when they are at their maximum.

The following chapters cover ideas that are applicable to response in general, regardless of the specific device involved. It is therefore awkward that different types of transducers have different units for response. This may be why there is no widely accepted custom for what algebraic symbol to use for response.

But there is a way around this difficulty. All physical signals transmit energy, so that one choice for measuring the "size" of a signal is to use power. As a signal passes from a radio station via the air to a radio antenna, or from amplifier through a wire to speaker, the rate of energy transmission is a power. This also means that for any transducer along the way, such as the radio receiver that is translating the radio signal into the electrical signal, the input and output of the transducer can both be measured by their powers as well.

Response measured in this way is called the **power gain,** often called simply **gain,** which in this book will be represented by g. Gain is a pure number, with no units. Suppose that the microphone mentioned above, under those specific conditions, gathers $W_{\text{in}} = 10^{-8}$ W of sound power, and the electrical signal out carries $W_{\text{out}} = 2.0$ μW. Then the gain is

$$g = \frac{W_{\text{out}}}{W_{\text{in}}} = \frac{2.0\mu\text{W}}{10^{-8}\ \text{W}} = 2 \times 10^{2} \quad . \qquad (88.2)$$

Notice that a device cannot violate conservation of energy. If power gain is greater than one, the extra energy in the output must be provided by some power source. Perhaps this microphone contains a battery, along with some circuitry that allows the battery energy to be added into the output signal. On the other hand, although the term "gain" sounds as if it implies an increase, in this technical sense a gain of less than one is entirely possible. It means that less power comes out than goes in. The lost energy does not vanish, but it has gone to some other purpose than representing the signal.

For amplifiers, since the input and output are the same type of signal, the signal amplitudes can be measured with the same unit, making the amplifier response unitless. However, there is still a choice of whether to measure the signal size by an amplitude-type characteristic or by power. Since amplitude is never proportional to power, these two types of response will not be equal; even more confusing, both types of unitless response might be called gain. However, this book will only use power gain. This avoids confusion and allows description of both amplifiers and transducers together at once. (If you have read Section 32b, you even know how amplitude gain and power gain are related mathematically.)

There is one drawback to focusing on power gain. For some types of signals, even when they represent a steady tone the energy may not flow smoothly. For instance, the energy may come in pulses, one pulse for each maximum and minimum in the signal versus time graph. It is therefore usually necessary to refer to the average power, averaged over at least one period. Most sounds have very short periods, so that this doesn't pose a huge difficulty. But it does mean that, unlike the amplitude response described at the beginning of this chapter, the power gain cannot be used to directly calculate the output oscillation from the input oscillation.

There are a few other names that response and gain go by in specific situations. When the transducer is designed to pick up very small inputs, as with microphones, the response might be called the **sensitivity.** In situations where there is no external energy supply, so that the output power comes from the input, then the gain might be called the **efficiency.** But other than having different connotations, these terms are synonyms.

Chapter 89. Response Curves

An ideal device produces an output signal that is a faithful copy of the input signal, the only potential change being an overall scaling by the response factor. (For transducers, that scaling includes a change of units.) But, as with most things in life, the ideal is never actually achieved in real systems. There are always limits, for instance in frequency or amplitude, which if exceeded cause the system to be non-ideal.

The least destructive way for a transducer or amplifier to be non-ideal is for it to remain a **linear device,** but to have a frequency dependent response. For the device to be **linear** means that for any possible signal, the spectrum of the input and the spectrum of the output have partials at the same frequencies. Each partial (with its specific frequency) passes through the device without changing its frequency, and unaffected by whether there are any other partials in the signal.

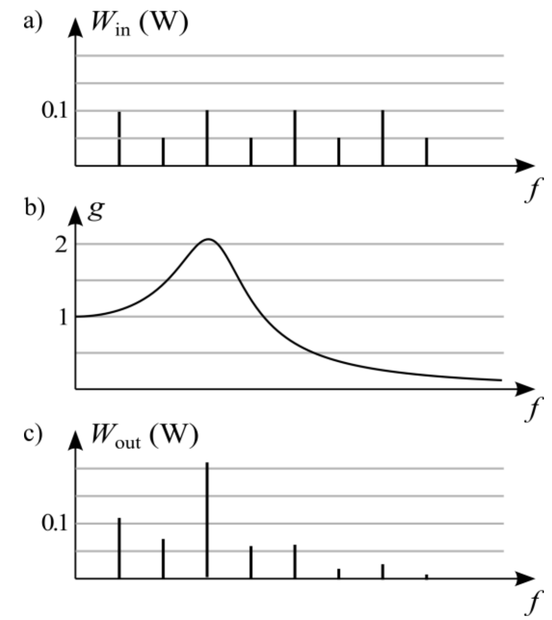

Figure 89.1

Applying a response curve to an input to calculate output.

Having a frequency dependent response means that Eq. 88.2 is modified so the gain is not a single value,

$$W_{out}(f) = g(f)W_{in}(f) \quad , \tag{89.1}$$

where the new aspect is that the equation is applied separately at each frequency. Figure 89.1 illustrates how this relates an input power spectrum to an output power spectrum. For each peak in the input spectrum, the gain *at that frequency* is used in Eq. 89.1 to determine the output peak height. The same idea could be applied to an amplitude response, but this book will only consider power gain.

A graph of response as a function of frequency is called a device's **frequency response curve.** Often either the first or last word of that name is dropped, and middle word could be replaced with **gain.**

This is often a very good model to describe a real device. If you have read Chapter 68, then you have already seen an example of a device that responds more at one frequency than at others, and perhaps you recognize Figure 89.1(b) as that device's response curve. It is a simple mass on a spring, but in that chapter the system was not considered as "transducing a signal."

The response curve in Figure 89.1(b) has a **resonance,** that is, a broad peak that is centered on a particular frequency. The response curve in Figure 89.2 is typical for a high-quality stereo speaker. It features a range of frequencies for which the gain is constant; within that range, the system has a **flat response curve,** so that it behaves like an ideal system. The gain isn't 1, but that doesn't disqualify this system as ideal. The output has a smaller power than the input ($g < 1$), but it can still be a faithful copy. Also notice that the horizontal axis of this graph is the logarithm of the frequency; this is very common for response curve graphs.

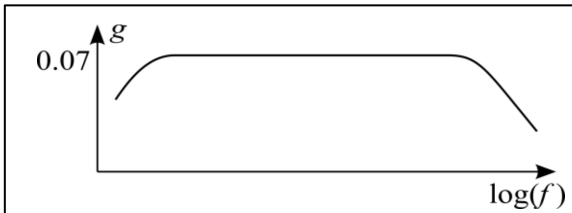

Figure 89.2

Example of frequency response curve of a device with a flat respone over a central frequency range, typical of audio equipment.

Chapter 90. Equalizers

Chapter 89 holds up a flat frequency response curve as "ideal." That is indeed the best behavior when the goal is to record or reproduce sounds, because we want the output to replicate the input. But there are other situations where that is not the goal. For example, Chapter 98 describes how the vocal tract uses non-flat response to produce the sounds in speech.

Even when a flat response is desired, it can be useful to have a non-flat response in one component in a system, in order to compensate for a non-flat response in another component. Suppose that you have speakers with poor low-frequency response. If you drive that speaker with an amplifier that intentionally boosts low frequencies, it is possible to get a final sound that matches the input sound signal.

Amplifiers that are specifically designed to enable control of their response curve are called **equalizers.** The name comes from their use to compensate for other devices in an audio system, thus making the combined response of the total system *equal* at all frequencies. Of course, equalizers can also be used to shape a sound signal according to the preferences of the user. **Graphic equalizers,** such as the one shown in Figure 90.1, have sliders on the front panel which even look like a response curve graph. If you look closely, you can see the frequency labels along the bottom, which roughly form a logarithmic axis.

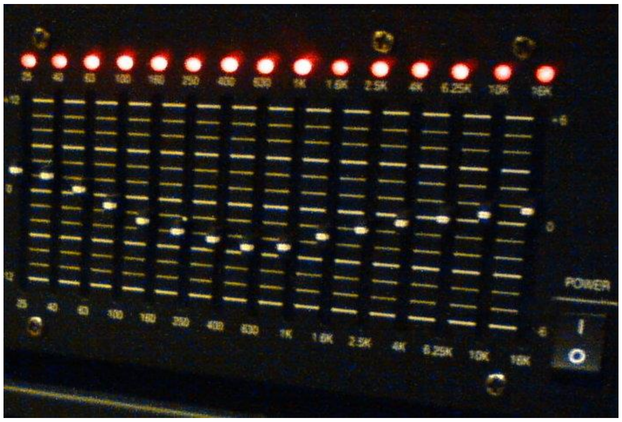

Figure 90.1

Graphic equalizer. Reprinted from César, *Wikimedia Commons*, 2008, https://commons.wikimedia.org/wiki/File:Graphic_equalizer.jpg (June 2015) / public domain.

An equalizer can successfully compensate for other devices in the system when those devices are all linear. For each partial in the signal that enters the system, the partial's frequency acts like an identification tag. That partial may get amplified or attenuated, relative to the rest of the signal, in one device, but it can be identified as it passes through a second device. Thus, its treatment in one device can be compensated for in another device.

To determine the equalizer settings (the gain at each frequency) necessary to compensate for the rest of the system, one could send pure tones with many different frequencies through the system, adjusting the gain for each one. For an equalizer with adjustments at many frequencies, this can be tedious. A slick way around this is to use a white noise input to the sound system. Recall from Chapter 45 that white noise contains power at all frequencies simultaneously. A skilled operator listening to the output can quickly recognize and correct for frequency regions that are amplified too much or too little.

Chapter 91. Sound Power Level

91a. SWL Defined

If you have read Chapter 57, you will notice that this subject very much parallels that chapter.

Because signals often involve very wide ranges of power, it is often helpful to graph power on a logarithmic axis. For this reason, and also because it simplifies certain calculations, we define a different way to specify power called **sound power level**, or **SWL**. "W" is used to abbreviate power because "P" is already used in SPL, where it stands for pressure. The usual algebraic symbol for SWL is L_W, and it is measured in units of decibels (abbreviation dB). The SI root unit here is the bel, and deci- is a metric prefix that is too

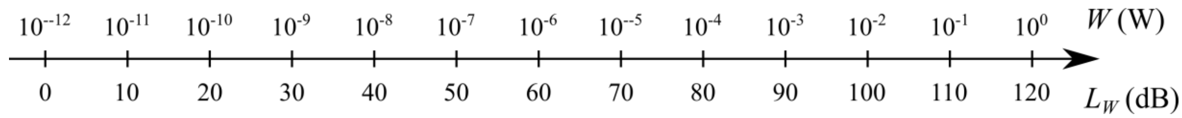

Figure 91.1

Axis showing relationship between Sound Power Level and power.

unimportant to make it into Table 4.1. But it is extremely uncommon to use anything but the combination dB.

Figure 91.1 illustrates how sound power levels are related to powers commonly encountered in audio applications. As one moves to the right along this scale, each tick mark corresponds to increasing the power by a *factor* of ten, while sound power level increases by an *increment* of 10 dB. Keep in mind that SWL and power are two ways of measuring the same thing, somewhat the way centimeters and meters are two ways to measure length. The conversion between them is more complex, though, using the equations[34]

$$L_W = (10\text{ dB})\log\left(\frac{W}{W_0}\right) \quad , \tag{91.1}$$

$$W = W_0 10^{(L_W/10\text{dB})} \tag{91.2}$$

$$W_0 = 10^{-12}\text{ W} \tag{91.3}$$

These equations can be thought of as the prescriptions for moving from one side to the other of the scale in Figure 91.1.

Each of the words in "SWL" means something specific. "Level" denotes something that has units of decibels and that is defined by an equation like Eq. 91.1. There are other "level" quantities, which differ from SWL only by what is inside the logarithm function; the word "Power" specifies that. The word "Sound" indicates not only the type of energy under consideration, but also what value to use for W_0. The value in Eq. 91.3 is chosen by convention; it does not have any physical significance.

Notice that zero plays a different role on the two scales. A power of zero would indicate no signal whatsoever. But because power is on a logarithmic scale in Figure 91.1, zero power doesn't even appear. If you try to calculate the SWL for $W = 0$ W, your calculator will probably give an error; based on the scale, we might call that $-\infty$ dB. On the other hand, an SWL of $L_W = 0$ dB doesn't mean no signal at all; that indicates a signal with power W_0.

91b. Extra: A Different Reference Level

The power of a signal representing sound, especially an electrical signal, is often expressed in decibels of a different sort. This is calculated using equations almost exactly like Eqs. 91.1 and 91.2. The only difference is that the reference power is not W_0, but instead

$$W_m = 1\text{ mW} \tag{91.4}$$

Following the naming pattern for other level quantities, this might be called **signal power level.** However, it does not have a widely accepted name, nor does it have a standard algebraic symbol. It does have a special unit. Levels measured with W_m as the reference level are given the unit dBm, which could be read as "decibels relative to one milliwatt."

[34] For the purist: see the footnote in Chapter 57.

Chapter 92. SWL Comparisons

92a. Comparing Two SWLs

If you have read Chapter 58, you will notice that this subject very much parallels that chapter.

Numerical comparisons can be made in two basic ways: differences (e.g., “My new car gets 5 more miles per gallon than the old one.”) and ratios (e.g., “My new car gets 1.2 times the mileage of the old one.”) Notice that difference comparisons require the choice of a unit (miles per gallon in the example), while ratio comparisons do not. One of the benefits offered by level measurements such as SWL is that they turn ratio comparisons into difference comparisons.

Suppose that we are comparing the power reaching the left and right speakers in a sound system. Comparing the SWLs as a difference, we find that

$$\begin{aligned} \Delta L_W = L_{WR} - L_{WL} &= (10\text{ dB})\log\left(\frac{W_R}{W_0}\right) - (10\text{ dB})\log\left(\frac{W_L}{W_0}\right) \\ &= (10\text{ dB})\left[\log\left(\frac{W_R}{W_0}\right) - \log\left(\frac{W_L}{W_0}\right)\right] \\ &= (10\text{ dB})\log\left(\frac{W_R/W_0}{W_L/W_0}\right) \quad , \end{aligned} \tag{92.1}$$

$$\Delta L_W = (10\text{ dB})\log\left(\frac{W_R}{W_L}\right) \quad , \tag{92.2}$$

where the special property of logarithms Eq. 55.3 has been used. The ratio of the powers translates to a difference in the SWLs. Notice how similar Eq. 92.2 is to Eq. 91.1. In fact, one might say that *all* SWLs are comparisons, with Eq. 91.1 really being a comparison to the standard W_0.

Level quantities are so directly tied to difference comparisons, that it is *never* appropriate to multiply or divide any quantity with decibel units. (The only exception is the operation with 10 dB when converting between decibels and their underlying quantity, as in Eqs. 91.1 or 91.2.)

When it is appropriate to compare or combine powers by addition or subtraction, however, SWLs become decidedly non-intuitive. Suppose that the powers in our example are $W_R = 0.02$ W and $W_L = 0.08$ W, so that $L_{WR} = 103$ dB and $L_{WL} = 109$ dB. This means that the amplifier driving the two speakers needs to supply a total power of

$$\begin{aligned} W_T &= 0.02\text{ W} + 0.08\text{ W} \\ &= 0.1\text{ W} \quad . \end{aligned} \tag{92.3}$$

The resulting total SWL is $L_{WT} = 110$ dB, a result which has no intuitive connection to the contributions of L_{WR} and L_{WL}. In such situations, it is safer to do calculations on the “power side” of the scale in Figure 91.1, rather than trying to work with decibels.

92b. Extra: Rules of Thumb

To avoid having to do calculations with Eqs. 91.1 or 92.2, it can be handy to have a few relationships memorized. Figure 91.1 clearly suggests than if one sound power is ten times larger than another (the separation between ticks), then the sound power levels for the two sounds differ by 10 dB. What may not be clear from that figure is that this relationship is true even if the two powers are not perfect powers of ten. A random example: powers of 83.5 mW and 8.35 mW also have SWL differing by 10 dB.

Similarly, whenever a power is doubled or halved, the SWL changes by almost exactly ± 3 dB. You can check this yourself using Eq. 92.2, by calculating $(10\text{ dB})\log\left(\frac{2}{1}\right)$ and $(10\text{ dB})\log\left(\frac{1}{2}\right)$. These examples display one benefit of working in decibels: the change in SWL is a simple numerical quantity, without needing to specify either the initial or final values.

Chapter 93. SWL Spectra

If you have read Chapter 62, you will notice that this subject very much parallels that chapter.

Considering Chapters 91 and 92, it will perhaps come as no surprise that the size of the partials in a power spectrum can also be measured in SWL. Equations 91.1–91.3 can be used to directly calculate either one from the other. Figure 93.1 shows an example of the same signal represented in both ways. A few features bear comment.

A relatively wide range of powers is compressed to a relatively small range of SWLs. This is one key aspect of level measurements (that is, measurements in dB). It's for this same reason that SWL is helpful in describing a wide range of powers. As a consequence, if the frequency axis in Figure 93.1(b) were positioned to cross the origin of the L_W axis, the vertical sizes of the four partials would be almost indistinguishable. For level measurements, it is usually a good idea to choose a range for the L_W axis that is not much larger than the range of peaks to be included. This means that the frequency axis will not cross at 0 dB; 0 dB won't even appear!

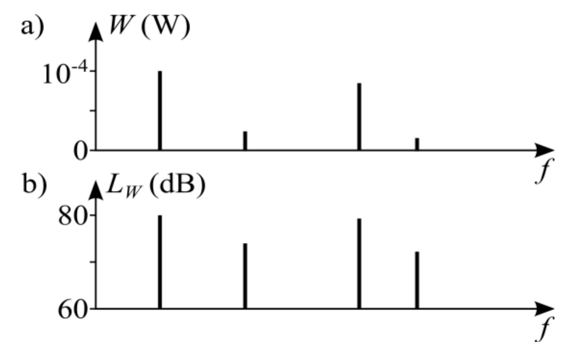

Figure 93.1
(a) Power spectrum. (b) SWL spectrum for the same sound.

Although Eqs. 91.1–91.3 are available to relate corresponding peaks (that is, a power peak and an SWL peak at the same frequency, both representing the same partial), it still may be helpful to calculate peak heights by using the relationships between partials at different frequencies. As a specific example, suppose that we focus on the relationship between the first two partials in Figure 93.1. Eq. 92.2 gives the relationship

$$L_{W2} - L_{W1} = (10\text{ dB})\log\left(\frac{W_2}{W_1}\right) \quad . \tag{93.1}$$

Just as with signals in general, the ratio of powers of partials gives the difference of the SWLs of those partials. In our specific example, the power ratio is

$$\frac{W_2}{W_1} = \frac{0.25 \times 10^{-4}}{1 \times 10^{-4}} = 0.25 \quad , \tag{93.2}$$

which directly gives the SIL difference in Figure 93.1(b)

$$L_{W2} - L_{W1} = (10\text{ dB})\log(0.25) = -6.02\text{ dB} \quad . \tag{93.3}$$

Chapter 94. Decibel Response

When measuring signal power with SWLs, it is handy to measure gain in a compatible way. By operating on Eq. 88.2, we can relate the input and output in terms of levels as follows:

$$\begin{aligned}\log\left(\frac{W_{\text{out}}}{W_0}\right) &= \log\left(\frac{gW_{\text{in}}}{W_0}\right) \\ &= \log(g) + \log\left(\frac{W_{\text{in}}}{W_0}\right) \quad ,\end{aligned} \tag{94.1}$$

$$(10\text{ dB})\log\left(\frac{W_{\text{out}}}{W_0}\right) = (10\text{ dB})\log\left(\frac{W_{\text{in}}}{W_0}\right) + (10\text{ dB})\log(g) \quad , \tag{94.2}$$

$$L_{W,\text{out}} = L_{W,\text{in}} + g[\text{dB}] \quad , \tag{94.3}$$

$$g[\text{dB}] \equiv (10\text{ dB})\log(g) \quad , \tag{94.4}$$

$$g \equiv 10^{g[\text{dB}]/10\text{ dB}} \quad . \tag{94.5}$$

Eq. 94.4 defines **gain in decibels** $g[\text{dB}]$. Normally, this is just referred to as **gain**, exactly the same term as used for g. In fact, even the notation [dB] is a special feature of this book. In common practice, the two types of gain are referred to by the same language, and it is the job of the reader to identify which kind of gain is intended. Eq. 94.5 simply reverses the definition, in case you need to obtain ratio gain from a known decibel gain.

The convenience of Eq. 94.3 is that when everything is expressed in decibels, only addition and subtraction are required for calculations, with no multiplication or division.

Chapter 95. Decibel Response Curves

If everything in Chapter 89 is expressed in terms of decibels, we again have a response curve that can be used to determine the output spectrum that will result from an input spectrum, as illustrated in Figure 95.1. As before, because we are assuming that the device is linear, each input partial results in an output partial at the same frequency. As before, the size of that partial is modified according to the value of the response curve at that same frequency. The new feature here is that the modification is by addition or subtraction, as given by Eq. 94.3. See if you can verify the relations in the figure.

When gain is expressed in decibels and frequency is on a logarithmic axis, a benefit arises in the response curves that might be unexpected. With those axes, sections of response curves often turn out to be made of straight lines. An example can be seen in the middle graph of Figure 95.1. That response curve is flat for higher frequencies and falls off in a line towards lower frequencies.

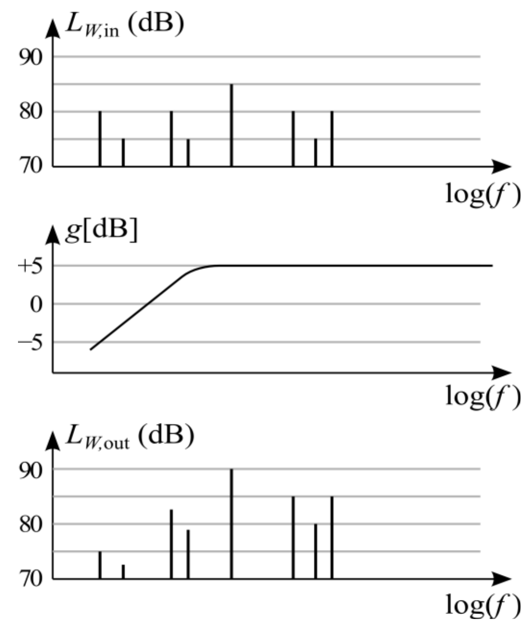

Figure 95.1

Applying a response curve to an input to calculate output, decibel version.

Chapter 96. Resonant Cavities

Chapter 28 describes how a volume of air with an opening or neck will oscillate at a preferred frequency, with the parts of the oscillation being a chunk of air moving in and out of the opening and the air in the volume compressing and expanding. Those two parts constitute a mass and a spring. Chapter 68 describes how, if such a system is driven at a variable frequency, the magnitude of the response depends on how close the driving frequency is to the natural frequency.

If a second small hole is now added, as shown in Figure 96.1 on the right side, that hole will emit sound originating from the compression and expansion of the internal air. This is the complete version of the Helmholtz resonator, employed so effectively by Helmholtz to analyze sounds in the late 1800s. In the days before electronics, it was extremely difficult to measure the vibrating displacement versus time associated with a sound, and quite impractical to make a sound spectrum through the FFT of the displacement versus time. To analyze a sound, one could listen to the sound through each of a set of Helmholtz resonators with different resonant frequencies.

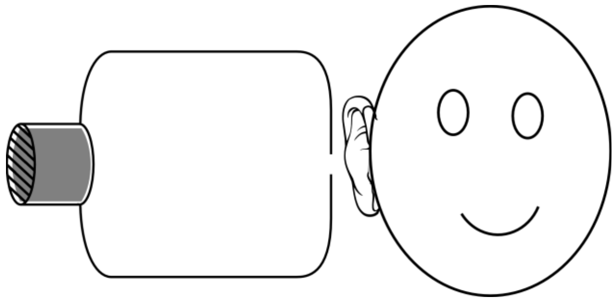

Figure 96.1
Helmholtz resonator used as frequency filter.

Now that the resonator has an input and an output, we have turned it into an amplifier. To be clear, recall that the label "amplifier" does *not* mean that the intensity of the sound is increased. The name applies because the input and output are of the same type, both being sound in air. The response curve can be defined precisely by Eq. 89.1, as for any other amplifier. Since the output is the direct result of the system oscillation, the response curve will look qualitatively like Figure 68.1; the height and narrowness of the response peak would depend on the Q of the resonator. If an input sound has a partial near the resonant frequency, then the output at that frequency would be emphasized compared to partials at other frequencies. So, by listening through a set of many resonators, the spectrum of a sound can be roughly determined.

The Helmholtz resonator is a particular example of a resonant cavity, a volume of air with a preferred oscillation frequency. In general, a resonant cavity might have multiple openings, and the motion of air in the natural oscillation might be more complicated. But that detail is not as important as the general fact that whatever sound enters the cavity, the partials and parts of the sound that are near a natural frequency are enhanced relative to other frequencies. Notice that the frequency of the input partial does not have to exactly match the resonant frequency. The resonance in the response curve can be quite broad (low Q), covering a range of frequencies.

Chapter 97. Vocal Tract

97a. Vocal Tract Structure

The human voice is created in the vocal tract, which is illustrated in Figure 97.1. Very similar structures apply for all mammals.

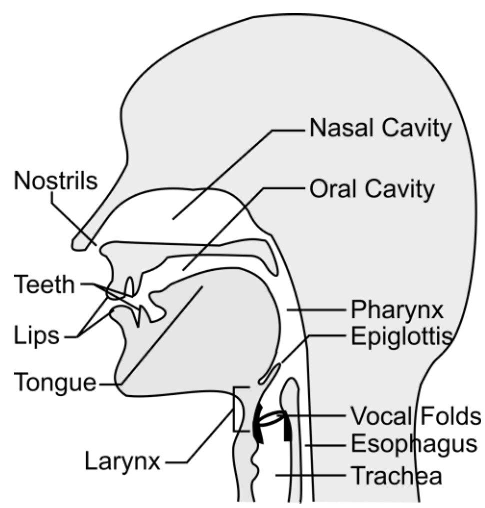

Figure 97.1
Parts of human vocal tract. Reprinted with added labels from Tavin, *Wikimedia Commons*, 2011, http://commons.wikimedia.org/wiki/File:VocalTract.svg (June 2015) / CC BY 3.0.

Voiced sounds originate at the **vocal folds** (sometimes called the **vocal cords,** although the term is somewhat misleading about their shape). These two membranes are located near the center of the **larynx.** The larynx sits just above the **trachea** (or **wind pipe**), which delivers air to and from the lungs below. The vocal folds can be closed, somewhat like lips, to control or stop this flow of air.

The **epiglottis** can close off the top of the larynx, but that is only used to ensure that food passes down the **esophagus** into the stomach. These parts do not normally play a role in sound production.

Sounds from the vocal folds pass up through the **pharynx**, and then either through the **nasal cavity** or the **oral cavity** (or both). The nasal cavity is relatively fixed in size and shape, as is its external opening at the

nostrils. The oral cavity is much more flexible, its size and shape depending on the position of the jaw and the **tongue,** and the size and shape of its external opening determined by the **teeth** and the **lips.**

97b. Vocal Folds

Phonemes are basic sounds, the fundamental parts that are combined to form language. Many vocal sounds, especially phonemes, involve rapid changes in the sound. This book will not address those sounds. We will consider only those vocal sounds that can be sustained. This essentially means the **vowel** sounds (pronounced with an open vocal tract), **fricative consonants** (pronounced with sound forced through a narrow opening in the vocal tract) such as *s* and *v*, and **nasal consonants** (pronounced with closed mouth) such as *m*.

For unvoiced fricative consonants, the sound production occurs at a narrow opening at some position above the vocal folds. For instance, between the tongue and roof of the mouth for *s*, or between the lower lip and upper teeth for *f*. The air rushing rapidly through the narrow opening has a chaotic, turbulent flow with rapid, random changes that create noise, meaning sound with a spectrum that includes no sharp peaks. The vocal folds can also operate in this manner, creating a narrow opening in the larynx, which is the origin of whispering.

In order to make different fricative consonants, or the different sounds required for articulation while whispering, the sounds must be different in some way. Although all these sounds are noise, they do differ in how much intensity they contain in broad ranges of frequency. This is partly accomplished by varying the shape and size of the narrow opening creating the sound. However, this does not account for all the variation, especially in the case of whispering.

The vocal folds can also operate in a second mode, which is responsible for voiced sounds including vowels, voiced fricatives such as *z* and *v*, and nasal consonants. The edges of the folds are held pulled tight and touching (or nearly so). The path for air through the larynx is cut off, but pressure from the lungs can push the vocal folds apart. The tension in the vocal fold edges provides a restoring force, which together with the mass of the vocal folds defines a natural frequency of vibration. This is certainly more complicated than the mass-on-a-spring arrangement discussed in Chapter 27, but following the general principles expressed in Eq. 27.2, the natural frequency can be increased by increasing the tension, which also slightly reduces the vibrating mass.

The air pushing its way between the vocal folds results in a vibrating motion. For soft to moderate sounds, the vocal folds actually are completely closed part of the time, preventing any airflow until sufficient air pressure builds in the trachea to blow open the vocal folds. When this happens, a puff of air is released. The velocity of this puff forms an almost triangular graph, as sketched in Figure 97.2. Pushing the air through more forcefully increases the size of the triangles, making them taller and wider until they touch. This does not, however, markedly change the frequency, which remains primarily determined by the features of the vocal folds themselves.

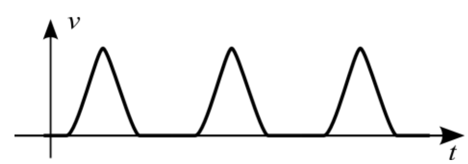

Figure 97.2
Air velocity coming out of larynx as a result of vibrating vocal folds.

Aside from the frequency and loudness, the shape of the oscillation that leaves the larynx cannot vary too much. So the question arises, how are a variety of vowel sounds made? The answer to this, as well as how a variety of noise-based sounds are made, lies in how the sounds change as they proceed out of the vocal tract. This is described in Chapter 98.

Chapter 98. Vocal Formants

The question of how to create the different sustainable phonemes of language is really a question of how to create different sound timbres, which are most easily described in terms of the shapes of their spectra, as in Chapter 42. A typical spectrum of the sound exiting the larynx, corresponding to the velocity graph in Figure 97.2, is shown in Figure 98.1(a). Since the velocity graph is periodic, the spectrum is harmonic (in this case, a relatively low pitch in the male vocal range). But more important for the current discussion is the overall pattern, which in this case is well described by the dotted line in Figure 98.1(a), called the **envelope** of the spectrum. The envelope does not itself describe a sound; rather, it describes a region of the graph that the actual spectrum stays within.

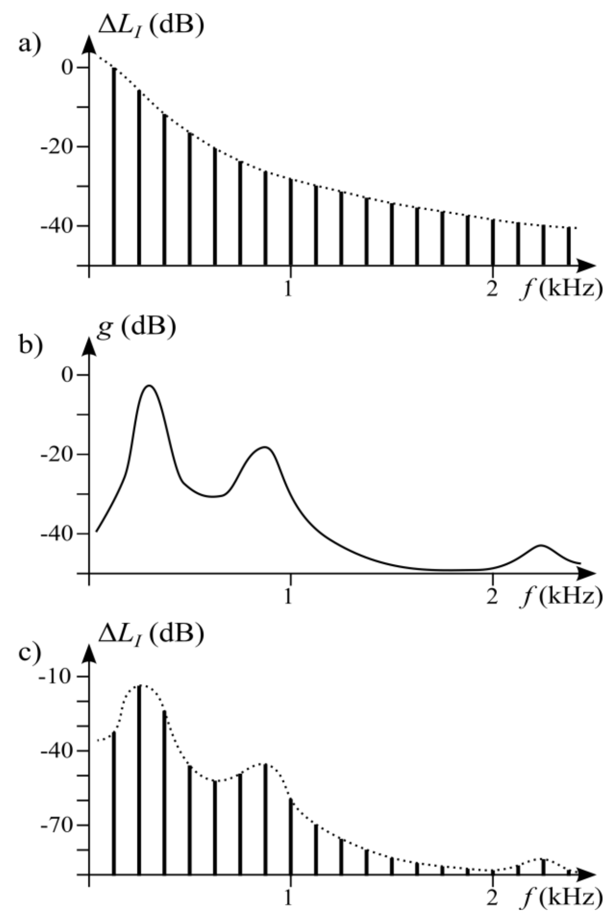

Figure 98.1

(a) Typical spectrum of voiced sound exiting the larynx, fundamental arbitrarily set to 0 dB.
(b) Response curve for ‘oo’ sound.
(c) Resulting vocal output. Note change in vertical scale.

As this sound passes through the vocal tract, it encounters three resonant cavities: the pharynx, the oral cavity, and the nasal cavity. These shape the sound with a response curve that features three broad resonant peaks, called **formants**. Figure 98.1(b) shows one example, the response curve for the ‘oo’ sound. It is tempting to try to associate each of the formants with one of the three resonant cavities. However, because those cavities are interconnected and strongly affect each other, the connections are not that simple.

The output of the whole system is found in the same way as for any other response curve. In Figure 98.1, the vertical axes are all in decibels, so they combine as in Chapter 95. The resulting output in Figure 98.1(c) continues to be harmonic, with the same fundamental as Figure 98.1(a), but the spectrum now stays within an envelope that has inherited some features of the response curve.

By adjusting the shape of our vocal tract (mostly the oral cavity), the frequencies and strengths of those three formants can be adjusted. The resulting change in timbre is what distinguishes the different phonemes. Table 98.1 gives data for a variety of vowels. The precise values of course depend on the individual speaking, but there are patterns, as there must be in order for us to understand one another. The strengths, expressed in dB relative to the strongest formant measured (formant 1 of ‘aw’), tend to be similar for all speakers. The center frequencies of the formants differ for men, women, and children, with men being lower and children being higher. But the frequency patterns are still similar. This is easier to see in Figure 98.2, which graphically represents the values from the table.

When one whispers, the initial sound coming from the larynx is noise, with no peaks in its spectrum. But that noise still passes through the same resonant cavities, which emphasize and de-emphasize frequencies in exactly the same way. This does not change the spectrum into one with peaks, but the noise sound exiting the system now has frequency ranges emphasized in the same way as voiced sounds. The result is that we can form recognizable phonemes, even without a harmonic spectrum.

Table 98.1

Formant frequencies and peak gains for vowels, averaged over 76 speakers. Frequency rows for (M)en, (W)omen, and (C)hildren. From Peterson and Barney (1952).

			phoneme									
	Formant		ee	i	e	ae	ah	aw	u̇	oo	u	er
center freq (Hz)	M	1	270	390	530	660	730	570	440	300	640	490
		2	2290	1990	1840	1720	1090	840	1020	870	1190	1350
		3	3010	2550	2480	2410	2440	2410	2240	2240	2390	1690
	W	1	310	430	610	860	850	590	470	370	760	500
		2	2790	2480	2330	2050	1220	920	1160	950	1400	1640
		3	3310	3070	2990	2850	2810	2710	2680	2670	2780	1960
	Ch	1	370	530	690	1010	1030	680	560	430	850	560
		2	3200	2730	2610	2320	1370	1060	1410	1170	1590	1820
		3	3730	3600	3570	3320	3170	3180	3310	3260	3360	2160
gain (dB)		1	−4	−3	−2	−1	−1	0	−1	−3	−1	−5
		2	−24	−23	−17	−12	−5	−7	−12	−19	−10	−15
		3	−28	−27	−24	−22	−28	−34	−34	−43	−27	−20

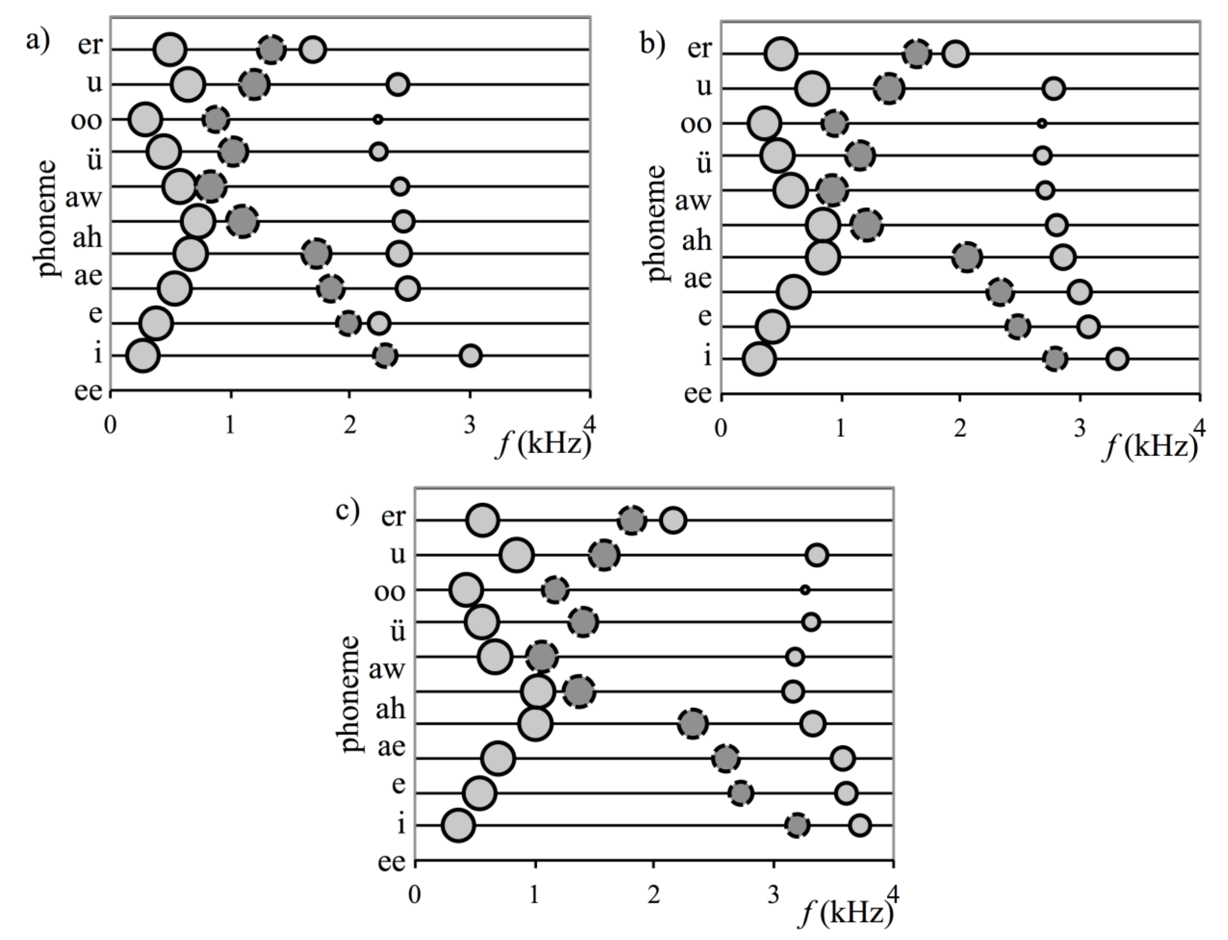

Figure 98.2

Formant frequencies and peak gains for vowels, same data as Table 98.1. Area of each circle is proportional to formant gain in dB above −46 dB. Second formant is represented in darker gray with a dashed border. (a) men, (b) women, (c) children.

In fact, using a spectrum of peaks to define timbre can be problematic when there are not enough peaks. When a soprano sings a high note, the fundamental might easily be as high as 625 Hz, higher than the first formant of many vowels. Figure 98.3 shows the same vowel as Figure 98.1(c), sung at a pitch of 625 Hz. (The formants are based on the values for males in Table 98.1, so this would be a falsetto voice.) The partials are drawn to the very same timbre envelope, but there are so few partials that it would not be possible for a listener to know what the envelope was. As a result, it is often difficult to distinguish between different vowels in high-pitched singing. This is not a limitation of the singing technique, but rather a limitation of the very mechanism by which phonemes are defined.

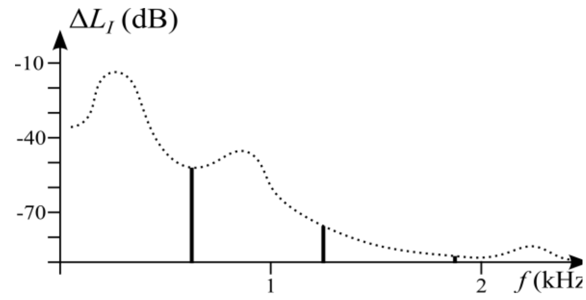

Figure 98.3
Same vowel as Figure 98.1, sung with fundamental at 625 Hz, high E♭.

Chapter 99. Damping

Chapter 33 introduces the idea of **damping** of a vibrating system. It can be thought of either as an external force that impedes the motion, or as a way that the energy is removed from the system (more details in Chapter 34).

The three graphs in Figure 99.1 illustrate the motion that results when an object with mass experiences both a restoring force (towards equilibrium at $\Delta x = 0$) and a damping force at three different strength levels. In all three graphs, the motion to the left of the dashed line is the SHM that results from having no damping force at all. At the dashed line, just when the SHM reaches maximum displacement, the damping force is suddenly activated. Or alternatively, you can imagine that the SHM was maintained by an external driving force that worked against the damping force, and that driving force is turned off at the dashed line.

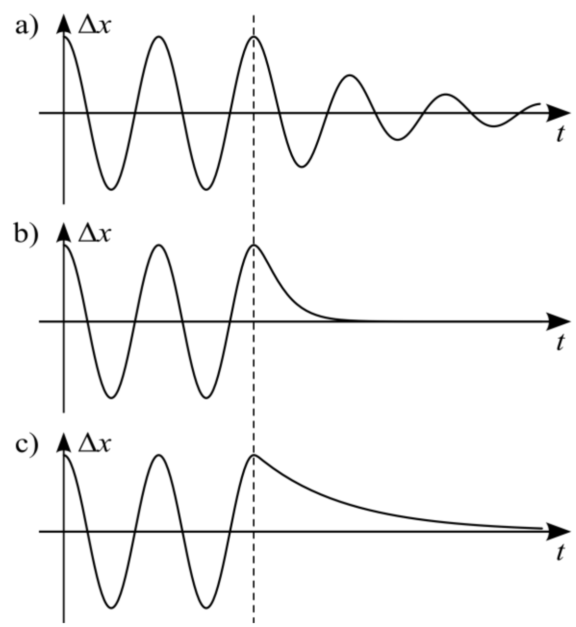

Figure 99.1
Different levels of damping. (a) Underdamping. (b) Critical damping. (c) Overdamping. Undamped motion to left of dashed line, damping turns on to right of dashed line.

In the examples of damping given in Chapter 33, the damping forces were weak enough to allow for oscillatory **damped harmonic motion.** A guitar string, which has very little damping, might die to half its initial sound intensity in 2 seconds; that roughly corresponds to loosing 0.1% of the amplitude with each vibration cycle. Figure 33.2 illustrates greater damping of about 2% of the amplitude per cycle, the damped oscillation in Figure 33.1 loses about 13% of its amplitude in each cycle, and Figure 99.1(a) loses 50% of its amplitude per cycle. That last one dies out much more quickly than the first, but those examples are all qualitatively similar. If the object is able to oscillate at all (without an external force shaking it), then the system is described as **underdamped.** For an underdamped oscillator, a weaker damping force results in the system taking longer to settle back towards the equilibrium state. The system gets back to the equilibrium *position* fairly rapidly (in a quarter cycle in Figure 99.1(a)) but it isn't in an equilibrium *state* because it has a high velocity.

If the damping forces are stronger, they may prevent free oscillations completely. In Figure 99.1(c), you might imagine that at the dashed line, a bucket of molasses is suddenly dumped on the oscillator. After that, the restoring force slowly pulls the object back towards the equilibrium position, but the damping forces from the thick molasses are so strong that the object can never pick up any speed, so that it can never get to the other side of the equilibrium position. A system like this is described as **overdamped.** For an overdamped oscillator, a stronger damping force results in the system taking longer to settle back towards the equilibrium state.

Both underdamped and overdamped describe wide ranges in the strength of the damping force. Where those two ranges meet, there is one special damping strength that results in a third type of motion, called **critically damped** and illustrated in Figure 99.1(b). This motion looks qualitatively like overdamped motion, since the freely moving mass never crosses through the equilibrium position. But the mathematical function of the critically damped graph is distinct from that of the overdamped graph. (The functions themselves are beyond the scope of this book.) One way to describe critical damping is that it is the damping force strength that results in the system returning to its equilibrium state in the shortest possible time.

Since a critical damping force gets an un-driven system back to equilibrium the most rapidly, it must remove energy from the system the most rapidly as well. Overdamping allows energy to stay for a while in the form of potential energy. Underdamping also allows energy to remain in the system, alternating between potential and kinetic forms as the system oscillates.

Chapter 100. Driven and Damped Oscillations

100a. Energy in Driven Vibrations

Chapter 31 describes that a natural, non-driven oscillation moves sinusoidally in such a way that its total energy (potential energy plus kinetic energy) is constant. A sinusoidally driven oscillator also moves sinusoidally (after any initial erratic motion has settled down). However, a driven oscillator does *not* have a constant total energy.

A driven oscillator still alternates between having potential energy in its spring and having kinetic energy in its moving mass. But because of the driving force, the maximum levels of those two energy types are no longer equal. Figure 100.1 shows how the maximum kinetic and potential energies depend on the driving frequency. While being driven at a particular frequency (that is, a particular value on the horizontal axis), the oscillator's total energy varies between these limits, moving up and down through the gray area of the graph. Only at resonance, where $f = f_0$, is the energy in the oscillator constant.

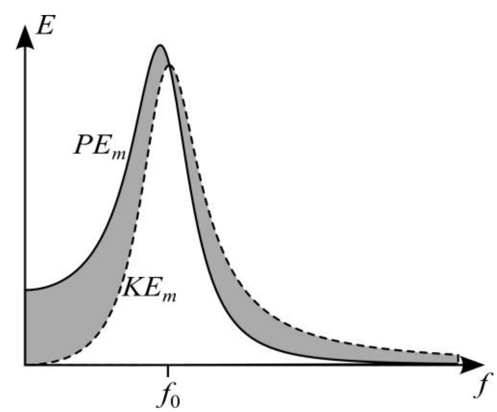

Figure 100.1

Energy range for mass on spring driven by sinusoidal force with frequency f. f_0 is system natural frequency and $Q = 2$. Solid line shows maximum potential energy, dashed line shows maximum kinetic energy.

This might seem to violate conservation of energy, but it doesn't. One reason that it doesn't is that work is continually being done between the oscillator and whatever external agent is applying the driving force. Sometimes, the external force is doing work on the oscillator, increasing its energy. Just as often, the oscillator is doing work on the external agent, so that energy is leaving the system. When the driving frequency is not close to the natural frequency, either significantly lower or higher, this exchange of energy is the main reason that the oscillator's energy varies.

Near the natural frequency, energy is leaving the oscillator in another way as well. The details of that are explored in Section 100b.

100b. Damping and the Response Curve

Since damping removes energy from an oscillator, in order for a damped oscillator to vibrate indefinitely, there needs to be an additional driving force putting energy into the oscillator. The Helmholtz resonator shown in Figure 96.1 is a good example. When an external sound reaches the large neck of the resonator, it tries to drive the resonator into vibration, putting energy in. As was described in Section 100a, if the driving frequency is not near the natural frequency, then this attempt is not successful; the frequency mismatch causes the vibrator to push the energy back out into the air. But if the driving frequency is close to the resonant frequency, then the energy does go into the resonator. As the energy accumulates, the resonator vibration amplitude gets larger and larger.

Meanwhile, a small amount of the vibration energy leaves through the small hole, in the form of sound that heads into the listener's ear. This departing energy constitutes damping of the vibration. It is not particularly easy to see where the damping force is, but a damping force must be there since energy is leaving. As the vibration of the resonator gets larger and larger, more and more energy leaves through damping. Eventually, the resonator vibration gets big enough that the rate of energy loss through damping equals the rate of energy gain from the driving force. This is the detailed picture leading to the resonance peak in Figure 68.1.

The result of all this is that the strength of the damping force is a factor in determining the height of the resonance peak in a response curve (that is, in the Q of the resonator). If the damping force is quite strong, then the resonator amplitude doesn't have to get terribly big for the energy flow to equalize. Thus the resonance peak will not be very tall, and Q will be smaller.

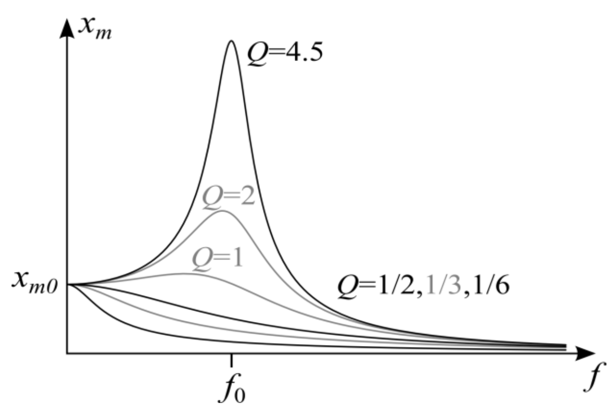

Figure 100.2

Resonance curves for damped driven oscillations. Black curves ($Q = 4.5, 1/2$, and $1/6$) correspond to damping levels shown in Figure 99.1.

The black resonance curves in Figure 100.2 correspond exactly to the three systems depicted in Figure 99.1. The underdamped example has $Q = 4.5$, so if it is driven it will show a distinct resonance. In fact all underdamped systems, which have $Q > 0.5$, have a peak in their response curve, although it is evident in Figure 100.2 that for roughly $Q < 1$ the peak is very small and no longer near the natural frequency f_0. The critically damped system, with $Q = 0.5$ has no resonance peak. Overdamped systems (Figure 99.1(c) has $Q = 1/6$) respond even less to a driving force.

To give a sense for the Q of less damped systems: Figure 33.1 has $Q = 26$, Figure 33.2 has $Q = 160$, and an acoustic guitar string has $Q \approx 3000$.

Chapter 101. Speakers and Microphones

Two transducers of particular importance for sound reproduction are speakers and microphones. Both of these translate between sound signals and electrical signals, making them **electroacoustic transducers**. Of course, they are each intended and designed to make that translation in a particular direction. But the physical coupling between electricity and motion is very similar in speakers and microphones. It is even possible to use a microphone as a speaker, and vice versa. They don't do a great job when used in the wrong role, but they do work. So, in the physics spirit of reducing things to the simplest possible terms, we will treat speakers and microphones together. As you read Chapters 102–108, be careful to notice when it is referring to speakers and when to microphones.

In both cases, the connection between sound and electricity is actually made by a third signal, a vibration. There are actually two devices involved, an **acoustomechanical transducer** to couple sound to a vibration, and an **electromechanical transducer** to couple vibration to an electrical signal. Chapter 102 and Chapters 103–108 respectively address these transducers in more detail.

The vibrating object at the center of all this has a displacement versus time which can be considered a signal representing the sound. In a logical sense, this object is between the two transducers, communicating the vibration from one transducer to the other. But practically, a single vibrating object is an element of both the electromechanical and the acoustomechanical transducer, so that the two transducers overlap.

The vibrating object must have a thin, flat part, such as a membrane or thin disk, which is called the **diaphragm**. The reasons for this shape are explained in Chapter 102. In speakers the diaphragm is often in the shape of, and therefore referred to as, a **cone.** Speaker cones generally have a very large opening angle, meaning that they are closer to flat disks than pointy arrowheads. They are also usually truncated cones, with the apex replaced by a small disk or dome shape.

Chapter 102. Acoustomechanical Transducers

Acoustomechanical transducers in microphones and speakers, which provide the coupling of vibration to sound, involve much less complexity than the electromechanical transducers. But they bring up a fundamental issue to be addressed.

Because air is a fluid, narrow or thin objects tend to cut through the air, with the air flowing around them. So, in order for the vibrating element of a microphone or speaker to interact strongly with the air, it needs to have a part that has a significant surface area and that vibrates in the direction perpendicular to that surface. This applies whether the transducer is trying to make the air move, or the transducer is trying to respond to the motion of the air. This surface is almost always called the **diaphragm**.

The diaphragm must be held in place somehow, so that it doesn't fall out of the transducer. But it must be able to move a little, so that it can vibrate. In most situations with a circular transducer, the part holding the diaphragm is called the **spider.** This name is inspired by the shapes that early versions of these supports took. As seen in Figure 102.1, they had many "legs," allowing the supported inner ring to move a bit. The majority of spiders now in use have replaced the "legs" with flexible materials, but the name has stuck.

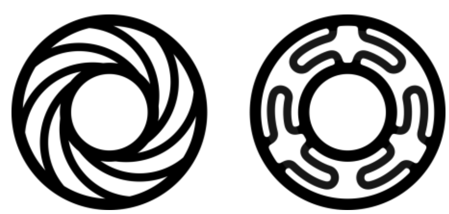

Figure 102.1
Examples of older-style spiders.

The vibrating part of the transducer must have some mass, simply because it is a material object. To hold it in place, it must also feel a restoring force, from a spider or similar part. From Chapter 23 we know that a mass experiencing a restoring force can oscillate on its own, with a natural frequency probably given by Eq. 27.2. And we know from Chapter 68 that an object with a natural frequency has a resonance at that frequency. All of which is to say that speakers and microphones *must* have a resonance peak in their frequency response curve. Yet that is the exact opposite of the desired behavior. The ideal transducer, in this case at least, is one with a perfectly flat response curve. What to do?

Another closely related problem is suggested by Figure 99.1. If one of these transducers is driven by an oscillation that suddenly stops, then we want the vibrating part of the transducer to stop as well. It would not be good if, when the sound signal to a speaker ends, the speaker continues to vibrate. Nor would it be acceptable for a microphone to continue to produce an electrical signal after a sound has stopped. This poor behavior is called **ringing.** It seems that we must avoid having an underdamped vibrator. On the other hand, when the driving signal stops, we do want the transducer to rapidly return to its equilibrium state, ready to respond to the next input. That makes an overdamped vibrator a poor choice. In order for the transducer to quickly and accurately respond when the sound signals stop, critical damping is the best choice.

Now returning to the issue of avoiding resonance peaks, we see from Figure 100.2 that critical damping solves that problem as well. A critically damped vibrator has $Q = 0.5$, which means that the resonance has been eliminated from the response curve. The response curve is not exactly flat, but at least it doesn't have a peak.

To get a response curve that is as flat as possible in the audio frequency range, it helps to make the natural frequency as high as possible. This suggests making the vibrator as low-mass as possible, and having a restoring force with a large spring constant. However, a large spring constant limits how much the vibrator moves in response to a signal. Figure 100.2 also shows that having the damping slightly underdamped, $Q \approx 1.0$, helps to flatten the response curve below the natural frequency.

The intrinsic need for quite a lot of damping explains the fact that speakers and microphones generally have quite low efficiency (also called power response), usually below 10%. In these days of energy conservation, it may seem wasteful for over 90% of the input energy to a speaker to be lost to damping. But that is the price we must pay in order to have transducers which accurately reproduce sounds.

Chapter 103. Electrical Basics

This book will go no deeper than a quick, general description of the most common types of electromechanical transducers found in speakers and microphones. The focus will be on the physical principles on which they operate, while practical details may be omitted. But for even cursory coverage, some elementary concepts of electricity must be introduced, in order to describe the 'electro-' side.

Electrical behavior starts with the fact that all substances are made of tiny particles (much, much smaller than microscopic) named **electrons, neutrons,** and **protons.** (You probably already know that these assemble to make atoms, although that is not particularly important for the current topic.) The electrons and protons have the property of **electric charge** with two different types, the type for protons being arbitrarily assigned as **positive charge** and the type for electrons assigned as **negative charge.** The two types attract each other very, very strongly, while charges of the same type repel each other with equal strength. The result of this strong attraction and repulsion is that in most objects that we encounter in everyday life, the two types of charge are mixed in exactly equal amounts, leaving the object **electrically neutral.** The two charges then effectively cancel each other, so that from the outside there is no evidence at all of the electrical nature of the parts inside. Even when we find an object that is "charged," it is really only a minuscule imbalance between the two charge types. This imbalance happens in cases of static electricity, but rarely in electric circuits.

Even when mankind puts electrical effects to use, the near-perfect balance between positive and negative charge remains in effect. For instance, in a battery, chemical reactions tend to push positive charges towards the positive terminal and negative charges towards the negative terminal. But the excess charges that accumulate at the two ends are far too few to detect through the forces that they cause. Instead, electrical circuits rely more on the motion of charged particles than on buildups of charge. The one significant exception is described in Chapter 104. (Many **electronic** circuits also utilize charge buildup, in tiny amounts on tiny parts, but these are not of interest for this book.)

When charge moves, it usually does so in the form of electrons moving through **metal**. In fact, it is the freedom of electrons to move inside metals that gives metals most of their characteristic properties (e.g., shiny, usually silver colored, good heat conductors). The flow of charges along a wire is **electrical current.**

Electrons move so easily through metals that even the undetectably small charge buildup on a battery's terminals can push enough current through a wire to have very detectable results, such as the glowing of an incandescent light bulb filament. At the other extreme, materials called **insulators** do not permit electrical currents at all. There are many materials between these extremes, through which current can flow, but not very easily.

Partly because the excess charges in electrical circuits are hard to measure, there is a more practical way to measure what causes currents to flow. It is called **electric potential** and is measured in the root unit **volt.** However, electric potential is not a force — it does not measure a push of any kind. To make an analogy, electric potential can be understood as a desirability rating for electrons. Every place in a circuit (in fact, every place in the universe) has a value of electric potential, which describes how much electrons "want" to be there (if we can imagine these particles to have free will). If two points in a circuit both have the same high potential and are connected by a wire, then there will be no current in the wire between them, because the electrons in one place don't "want" to go to the other, equally nice place. In order to get a current, you need to connect a place with high potential to a place with low potential. Then a current will flow, as the electrons flock from the low potential to the high. Thus, currents result from **electric potential differences,** also called **voltages.** (Physical relationships do exist between potential differences and the forces, measured in newtons, that push electrons. But these relationships are pretty much useless in understanding electrical circuits.)

Since the electrons repel each other, they don't like to get too crowded. Excess positive charge will result in a higher electric potential, since the positive charge attracts the electrons. But as electrons arrive, the potential is lowered, and any current quickly stops. This feature, that the motion of some electrons changes how quickly other electrons will follow, is often at the core of what gives students difficulty in studying electricity. In order to have a sustained current, one needs something like a battery to continually maintain the potential differences by supplying a little excess charge. The battery acts as a charge pump: not a reservoir of charge, but a way to push charge around.

Out of the physical quantities of charge, current, and voltage, it turns out that voltage is the easiest to measure and control. Electrical signals representing sound are most often voltage varying with time. Sometimes, especially in the context of electromechanical transducers, it is easier to understand the operation in terms of a current signal. It is fairly easy for an electronic circuit to convert a current signal into a voltage signal, or vice versa, but the details are beyond the scope of this book. In the transducers to be described here, sound is *never* represented by a charge that varies in time.

Chapter 104. Condenser and Electrostatic Transducers

Condenser type microphones are based on an electrical component called a **capacitor**. In the early days of electronic circuits, capacitors were also called condensers. Today, however, the term condenser is very rarely used except for microphones.

A capacitor is made of two metal pieces brought very close to each other without touching. Usually the pieces are thin **plates** with a large area, as shown edge-on in Figure 104.1. This allows a capacitor to break the usual rule of having no significant excess charge. If excess positive charge is put on one plate, and excess negative charge is put on the other plate, then their strong attraction to each other, across the small gap, helps to hold them there. Charge balance is still enforced in a sense, because the magnitudes of positive and negative charge on the two plates are never detectably different from each other.

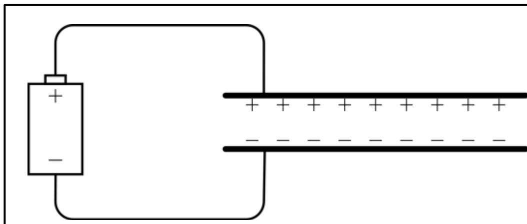

Figure 104.1

Capacitor (two close metal pieces on the right) being charged by battery.

The attraction across the gap is not strong enough by itself to keep the charges in place. If a wire is connected between the two sides, the excess electrons on one side will rapidly move to the other side, canceling the positive charge and leaving every part electrically neutral. This is called **discharging** the capacitor (even though the total charge on the whole capacitor is always zero!). But with the help of attraction across the gap, a battery is able to segregate significant amounts of charge. In fact, in situations like Figure 104.1, the amount of charge on each plate is proportional to the voltage applied by the battery.

Instead of a wire, the two sides could be connected by an object through which it is difficult, but not impossible, for current to flow. Such an object is called a **resistor.** This would still allow the capacitor to discharge completely, but it slows the process down. By choosing how much the resistor impedes the current flow, a circuit designer can adjust how long the discharge takes, from nanoseconds to minutes.

In order to function, a condenser microphone must have a voltage source to **bias** the capacitor, which means providing an initial potential difference and hence an initial charge separation. In Figure 104.2 that source is shown as a battery in the microphone, but very often the potential difference is supplied from an external source through the microphone cable.

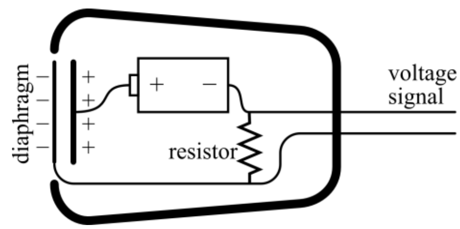

Figure 104.2

Schematic of basic operational elements of condenser microphone, including battery as source of bias voltage.

In a regular capacitor the gap between the plates is filled with a rigid material. But to make a microphone, the gap is not filled, and one of the "plates" is a flexible foil that forms the diaphragm, on the left in Figure 104.2. When a sound vibrates the diaphragm, the gap varies, getting smaller and larger. When the gap gets smaller, the attraction between the charges gets stronger, and the battery would normally push more charge on. However, there is a resistor in the way (the zig-zag line in Figure 104.2), chosen so that in one period of a sound signal, hardly any charge has time to flow through to the capacitor. The result is that the charge on the diaphragm is nearly constant, and all that changes is the *desirability* of adding electrons there— that is, the electrical potential of the diaphragm.

Figure 104.2 implies that the resulting voltage signal leaves the microphone directly, as a voltage (i.e., a potential difference) between the wires to the right. But the majority of condenser microphones include a built-in preamplifier, which increases the amplitude of the voltage signal. Often that preamplifier is powered by the same voltage source that is used to bias the capacitor/diaphragm.

The need for a biasing voltage is avoided by an innovation from the early 1960s: the addition of an **electret** to make an **electret condenser microphone.** [35] An electret is a manufactured insulating thin film, usually a plastic, which has permanent excess charge on its surfaces, positive on one side and negative on the other. It is like a frozen charged capacitor. The electret is placed between the condenser plates, usually either making part of the diaphragm (as in Figure 104.3) or adhering to the back plate. The electret's permanent charges then attract corresponding charges onto the condenser plates, effectively biasing it without a power source. From here the operation is essentially the same: because the charge on the electret cannot change, the charge on the diaphragm (and the back plate) is also nearly constant, but a vibration of the diaphragm causes a voltage signal.

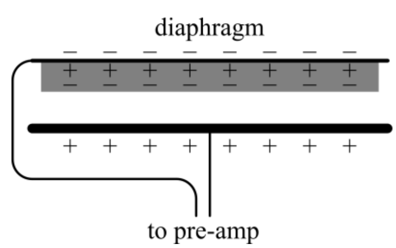

Figure 104.3

Wlectret (gray rectangle) added to inner surface of diaphragm of condenser microphone.

[35] G. M. Sessler and J. E. West, "Self-Biased Condenser Microphone with High Capacitance." *J. Acoust. Soc. Am.* 34(11) (1962): 1787-1788.

Electrostatic speakers are also based on a capacitor with one plate that is flexible. Cause and effect are reversed, so that varying amounts of charge change the attraction between the capacitor plates, and the flexible plate moves in response. To make a rudimentary electrostatic speaker, the capacitor in Figure 104.2 (including the flexible diaphragm) could be connected directly to two wires with a voltage signal between them. A varying voltage on the wires would vary how much charge is pushed onto the two capacitor plates (although they always maintain balanced charges). More charge on the plates results in a stronger attraction between them, so that the diaphragm flexes in. Less charge allows the diaphragm to relax outward. The diaphragm can never be repelled; reversing the charges (swapping the positive and negative) would still cause an attraction. Therefore, the voltage signal would need to be a special one that always has one wire at a higher electrical potential than the other.

The design in the previous paragraph relies on tension in the diaphragm to counteract the attraction of the charges. A better configuration puts two capacitors back-to-back, as in Figure 104.4, so that electrostatic forces can attract in both directions. A voltage source **biases** the central diaphragm relative to the top and bottom plates. (The voltage source is represented by a battery in the figure, but actually a high voltage power supply is needed.) While the speaker is not making any sound, the electrostatic attractions of the diaphragm towards the two outer plates cancel each other.

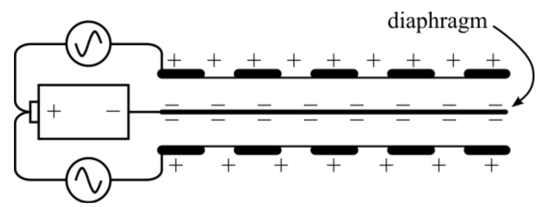

Figure 104.4

Symmetric electrostatic speaker. Top and bottom plates are dashed to represent holes, allowing vibrating diaphragm to deliver sound through them. Note that top and bottom plates have slightly different charge, although together they balance charge on diaphragm.

To make sound, a voltage signal is then added to one plate, while its inverse is added to the other plate. The circles showing one sinusoidal cycle represent devices that vary the voltage between their ends according to the audio signal. At the moment depicted in Figure 104.4, the upper plate has a little extra positive charge and the bottom plate is a little short on positive charge, although the total charge in the picture is still neutral. The resulting upward force on the diaphragm will cause it to flex upward. Rapidly varying the voltage will make the diaphragm vibrate. (It is not really part of the electromechanical behavior, but notice that the top and bottom plates need to be perforated by many holes, so that the sound produced at the diaphragm can get out.)

Condenser microphones are generally considered the most accurate type available, being quite sensitive and with excellent high-frequency response because of the very low mass of the diaphragm. They can also be made quite small, with dimensions less than a centimeter. However, they are not the most rugged microphone type, and they do require external power. Even the electret types require power, because they still need the preamplifier. Electret-based microphones tend to be of lower quality, but not due to the physics of their operation; it’s just that they can be cheaply produced.

Electrostatic speakers are also very accurate, with a very flat frequency response and very little distortion. Because they are shaped like panels, they can create better ambient sound, but because the panels project sound forward and backwards but not to the sides, they can be difficult to position in a room. These speakers also usually have poor performance at very low bass, partly because they cannot use enclosures to emphasize it (see Chapter 175). Finally, they require both external power and special circuitry to connect to standard sound amplifiers. As a result, they are not the most common type of speaker.

Chapter 105. Piezoelectric Transducers

One common electromechanical transducer is the **crystal,** or **ceramic,** transducer. This is based on **piezoelectric** materials, which have the property that they change shape slightly when a voltage is applied between two sides. One of the first known piezoelectric materials was quartz crystal. This is why microphones based on these transducers are called “crystal” microphones. However, the most common piezoelectric materials used today are ceramics, not crystals. Speakers based on this type of transducer are much more likely to be called piezoelectric or piezo speakers.

In Figure 105.1, the long gray rectangle on the right represents looking edge-on at a thin plate-shaped piece of piezoelectric material. The top and bottom surfaces of the plate are coated with metal electrodes (the heavy lines). When a battery (in the dotted ellipse in the middle) is connected to those electrodes with wires, causing a voltage between the electrodes, the piezo responds by getting thinner and wider, as represented by the dashed-line rectangle. If the battery were reversed, the plate would get thicker and less wide.

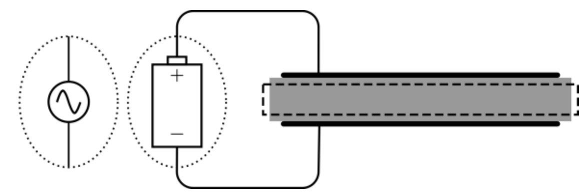

Figure 105.1

Schematic illustrating piezoelectric behavior. Piezo material (right) changes shape due to voltage from the battery or oscillator (left).

The shape change is tiny. A 1.5 V battery would typically change the thickness by less than 10^{-6} mm, which is far too small to have a significant effect on the surrounding air. In order to cause larger motion, it is necessary to **mechanically amplify** the motion. Figure 105.2 shows one way to achieve this. Two piezo pieces, shown with different shades of gray, are bonded together so that they have opposite behavior. When one expands the other contracts, and that causes the pair to bend as shown by the dashed lines. This allows for small but macroscopic motion of the diaphragm, although the deflection in Figure 105.2 is exaggerated for clarity.

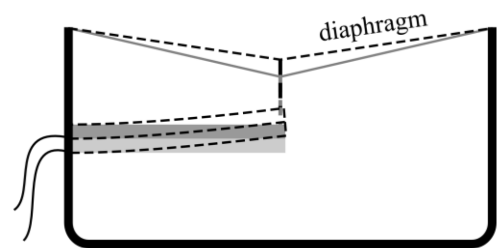

Figure 105.2

Basic construction of piezo speaker or crystal microphone. Pair of piezo elements (two shades of gray) allows moderate voltages to cause significant deflection.

Returning to Figure 105.1, if we replace the battery with a time-varying voltage source (represented by the symbol in the dotted ellipse to the left), for instance one varying with a signal that represents a sound, then the piezo will move to replicate the sound, and we have a piezo speaker. The piezoelectric effect works in the opposite sense as well: if the piezo element is compressed or stretched by an outside force, the piezo element creates its own voltage difference between the electrodes. If the voltage sources in Figure 105.1 are completely removed, and we arrange for the air pressures of sound to distort the piezo element, then a small time-varying potential difference signal will be created between the wires. This is the fundamental basis for a crystal microphone. Notice that neither the microphone nor the speaker requires electrical power, beyond the input electrical sound signal for the speaker.

One advantage of this transducer is that because it is made of a solid material with few moving parts, it can be quite rugged. For similar reasons they can be inexpensively produced and can be quite small. Indeed, it would be difficult to make a piezoelectric transducer larger than a few centimeters. So, for instance, piezoelectric speakers are ideal for hiding inside greeting cards.

However, in both modes of operation there are drawbacks in the frequency response. Crystal microphones generally have poor response for the highest frequencies. This is primarily because of the mass which must be accelerated, so that it is difficult for rapid sound pressure variations to move the piezo.[36] On the other hand, piezoelectric speakers generally have poor response for low frequency sound. This is primarily because they cannot be very large. There are several reasons that the best speakers for bass pitches need to be large, some of which are described in Chapters 126 and 172.

[36] S.P. Bali, *Consumer Electronics* (Singapore: Pearson Education, 2005), 28.

Chapter 106. Electromagnet Basics

Voltage is not the only way to cause a current. If a **magnet** moves relative to a wire, that can also push a current through the wire. This book will not delve into the concepts that are used to understand this effect. It is enough to know that magnets have two ends, called **poles** and conventionally named **north** and **south.** Generally, in order to cause current flow the poles of the magnet need to move in a direction perpendicular to the length of the wire. A stationary magnet has no effect, and moving the magnet along the wire has no effect. The magnet is not attracting or repelling the electrons. Nor is the magnet creating a potential difference. Rather, the motion and the magnet's properties combine to push the current directly.

Like many of these electromechanical effects, this one can be reversed. If a magnet is near a wire, and a current passes through the wire (perhaps caused by a battery between the distant ends of the wire), a resulting force between the wire and the magnet will try to cause motion. Again, the force will generally be perpendicular to the wire, although the details depend on the shape and arrangement of the two parts.

This connection between magnets, electrical currents, force, and motion is named **induction.** For instance, a current resulting from a moving magnet is called an **induced current.**

Chapter 107. Ribbon Microphone

The **ribbon microphone,** which was especially popular in the early days of radio, is based very directly on the induced current effect. Figure 107.1 shows the basic elements: a very thin metallic ribbon (corrugated for stiffness) exposed to the air is suspended between the poles of a C-shaped magnet. When a sound causes the ribbon to vibrate (in the figure, in and out of the page), an induced current flows left and right in the ribbon, thus causing a current signal in the wire that is attached to the ribbon ends.

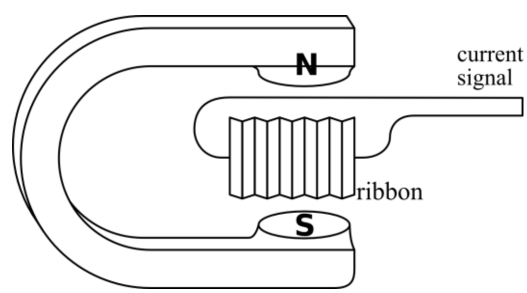

Figure 107.1
Schematic of essential parts of ribbon microphone.

The very light ribbon with very little restoring force mostly responds directly to air motion, as opposed to responding to the forces of the air pushing on the ribbon. Because of that, it is not technically a **diaphragm,** although the distinction is a fine point. Because it responds to air motion, and because the signal occurs whenever the ribbon is moving, this microphone is also called a **velocity microphone.**

Ribbon microphones are primarily sensitive to sounds that arrive perpendicular to the ribbon surface, while sounds from other directions are not detected as well. This can be either useful or a nuisance, in different situations. These microphones are not very sensitive, and don't have the best high frequency response. They are also somewhat delicate, so that they can be damaged by very loud sounds.

There is no practical speaker that is a direct analog of the ribbon microphone, because the induction effect is very weak. In the microphone, when a sound causes a very small current signal, we can effectively **electronically amplify** the signal. But if a current signal was applied to the ribbon, the resulting tiny induced motion would not make an effective speaker. Unlike the piezo speaker in Chapter 105, this motion cannot be mechanically amplified because the ribbon has to be so thin and flimsy; if it were made heavier, then it would move even less in response to the current. One could imagine amplifying a current signal before running it through the ribbon, but a current large enough to move the heavier ribbon would also melt the ribbon!

Chapter 108. Dynamic Transducers

Dynamic speakers and microphones (also called **moving coil** speakers and microphones) use the same induction effect as the ribbon microphone. They get around the smallness of the induction effect by using a **voice coil,** in which wire winds in a circle hundreds of times. Each turn in the coil participates in the induction, strengthening the total effect, but the parts need to be completely rearranged compared to the ribbon microphone.

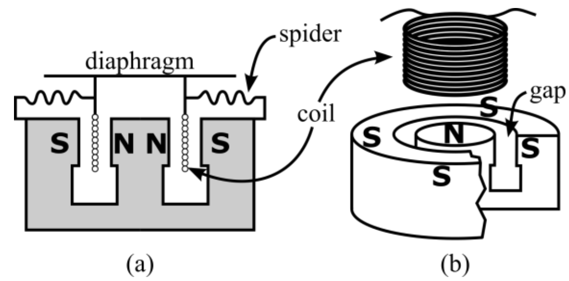

Figure 108.1

Schematics of essential electromechanical parts of dynamic speaker or microphone. (a) Cross section through device, with artificially simple diaphragm. (b) Coil and magnet, part cut away to reveal internal shape.

The magnet is shaped as shown in Figure 108.1(b). A pie-wedge cut from this cylinder would resemble the magnet in Figure 107.1, but the outer "pole" of this magnet makes a complete ring. The coil fits into the gap between the magnet's poles, which is made as narrow as possible so that the magnet's interaction with the coil is maximized. Figure 108.1(a) shows how a spider holds the coil in the gap so that it is able to move along its axis without rubbing.

The induction effect works from relative motion, so either the coil or the magnet could be the moving part. Early **moving iron speakers** held the coil stationary (and also had a different magnet shape). However, the coil can be made less massive, and thus easier to move, than the magnet. Therefore, in modern dynamic devices the magnet is held stationary and the diaphragm is attached to the coil.

As indicated in Figure 108.1(b), the two ends of the wire that forms the voice coil are not connected inside the transducer. For a dynamic microphone, the wire ends extend elsewhere and are connected. Vibration of the diaphragm, and thus of the coil, causes an induced current signal in the wire, which is detected at the far ends of the wire. For a dynamic speaker, a current signal is driven in the coil wire (perhaps by a voltage signal applied to the two ends), which by induction causes an oscillating force on the voice coil.

Dynamic microphones are nearly as common as condenser microphones. Although the dynamic type is not as sensitive and doesn't have as good high frequency response (due to the larger mass attached to the diaphragm), it doesn't require a power source. The dynamic design has more pieces than the condenser, but they can be made more sturdy and less sensitive to environmental conditions. More pieces does mean that the microphone cannot be as small, the smallest being handheld size.

Dynamic speakers, however, are by far the most common sound system speakers. They can most easily be designed for large amplitude motion of large diaphragms that are required to generate good low frequency response (see Chapters 126 and 172). These **woofers** require relatively large coils, the mass of which (together with the large diaphragm) is difficult to accelerate fast enough to generate higher frequencies. As a result, such a speaker will not have good response across the entire audible frequency range. But this is easily addressed by combining multiple speakers into a speaker system. A small **tweeter,** with a smaller and lighter coil and diaphragm, handles the high frequencies. For even better overall frequency response, three or more speakers can be combined, with each one responsible for a specific frequency range.

Chapter 109. Distortion

Any change to a sound signal that involves more than a non-flat response curve is called **distortion.** Distortion causes partials in the output signal at frequencies that were not present in the input signal. Generally, this is much more difficult to compensate for than a non-flat response curve. The problem is that for each partial in the output signal, it is difficult to know whether that partial should be there, having come from the original input signal, or whether that partial arose as a result of the distortion.

Distortion is not necessarily a bad thing. Various types of distortion are often deliberately added to electric guitar performance. Also, distortion is an intrinsic part of non-proportional signals, which as mentioned in Chapter 88 can be useful under controlled conditions, although this book will not explore that area.

Distortion tends to arise for large signals and loud sounds, when the devices involved are pushed the farthest from the equilibrium that they oscillate about. Although distortion is of particular interest in the field of sound reproduction, it can happen with any transducer or amplifier. Even our own ears can cause distortion when listening to loud sounds.

Chapter 110. Harmonic Distortion

The most common type of distortion is **harmonic distortion.** This is distortion which creates partials at harmonics of the input partials. That is, if the input signal has a partial at frequency f, the output signal will have a partial at f, but also at some multiples of f ($2f, 3f$, etc.). Figure 110.1 shows an example, where each input partial results in additional partials at the second and third harmonics.

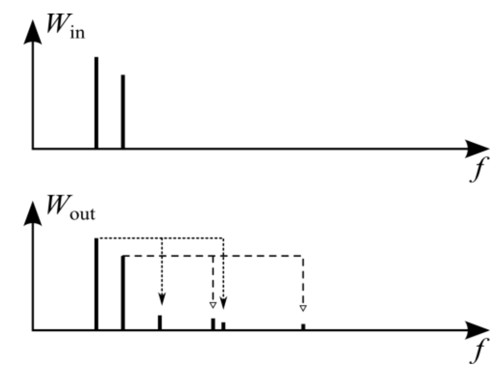

Figure 110.1

Example of harmonic distortion.

One way to cause harmonic distortion is **overdriving,** in which the signals have amplitudes larger than the device is designed to handle. This can distort the signal in various ways. One simple example is **clipping,** which is illustrated in Figure 110.2 for an output voltage signal from an amplifier. Here the pure tone input times the gain should result in the output indicated by the dashed line. However, the amplifier is unable to produce voltages outside the range $\pm V_{max}$, with the result that the actual output follows the solid line, which is clipped off at the top and bottom.

In Figure 110.2, since the shape of the actual signal is not a sinusoid, we know that the spectrum must have multiple partials, compared to one partial for the input. But since the signal is still periodic, with the same frequency as before, we know from Section 44a that the resulting output spectrum must be harmonic. Thus, we know that this situation will result in harmonic distortion. It would take some Fourier analysis work to figure out numerically how large the added partials would be.

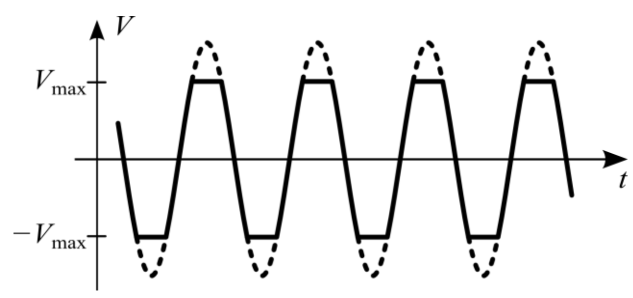

Figure 110.2

Example of signal clipping due to overdriving. Dashed line is the ideal ouput, solid line is actual ouptut.

For high-fidelity sound reproduction and other applications where distortion should be avoided, a figure of merit often used is the **total harmonic distortion.** With a pure tone input, the total harmonic distortion is defined as the square root of the ratio of the power in all the output overtones to the power in the output fundamental,

$$\text{THD} = \sqrt{\frac{W_2 + W_3 + W_4 + \cdots}{W_1}} \quad , \qquad (110.1)$$

where the subscripts indicate the harmonic numbers of the partials. Why the square root? Considering proportion 32.1, the square root is to make THD relate more closely to amplitudes. In more detail, the overtones must be combined by adding their powers together, as opposed to adding amplitudes, because conservation of energy ensures that is a meaningful operation. But then, to relate the metric to amplitudes, a square root is taken.

THD is sometimes given in decibels instead,

$$\mathrm{THD[dB]} = (10\ \mathrm{dB})\log(\mathrm{THD}^2) = (20\ \mathrm{dB})\log(\mathrm{THD}) \quad . \qquad (110.2)$$

The THD is squared in this equation so that the decibels are calculated with respect to power, which is consistent with our definitions of power level and decibel gain in preceding chapters. In the simplified form of the last member of Eq. 110.2, this gives rise to the 20 dB multiplier. This may be a surprise, but it actually shows up whenever decibels are calculated based on amplitudes. This book has been careful to avoid such calculations, to avoid confusion.

One value of total harmonic distortion cannot completely describe how a device behaves, because THD depends on the frequency and amplitude of the pure tone input, and also on the gain (for devices that have variable gain). Devices are therefore characterized by their **maximum THD** over a range of conditions. When comparing two devices to see which is better, it is important to know whether they were tested over the same range of conditions.

Chapter 111. Intermodulation Distortion

There are many other kinds of distortion besides harmonic distortion. This book will describe just one other, **intermodulation distortion.** This one is interesting because it is not apparent with a pure tone input, but only shows up when the input has multiple partials. It is true, though, that a device which exhibits intermodulation distortion is likely to also exhibit harmonic distortion for a large amplitude pure tone input.

Intermodulation is easiest to see with an input consisting of two partials, one at a much lower frequency than the other, $f_L < f_H$. As illustrated in Figure 111.1, the output has additional partials at frequencies that differ from the higher frequency by a multiple of the lower frequency,

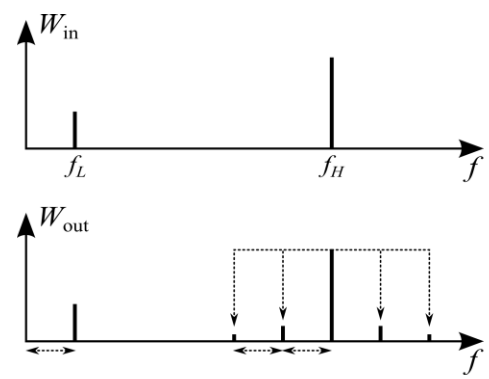

Figure 111.1
Example of intermodulation distortion.

$$f_{\mathrm{IMD}} = f_H \pm n f_L \quad , \qquad (111.1)$$

where n is any integer. The size of the distortion peaks are normally smaller for larger values of n.

In a musical setting, intermodulation distortion is considered to be more harmful than harmonic distortion. Musical sounds already are composed of harmonic partials, so when harmonic distortion adds to those partials at harmonically related frequencies, they tend to blend in with the original pattern. This will change the timbre of the sound, but not the pitches. In contrast, the frequencies that result from intermodulation distortion are likely not harmonically related to any of the input frequencies, so they are much more noticeable.