Chapter 7

Sound recording and reproduction

The vast majority of the music we listen to today is recorded, and recording technology as a result plays a substantial role in the ecosystem of musical sound and performance. With the exception of the very earliest recording technologies over a century ago, all modern recording works in essentially the same way: sound is converted into an electrical signal, which is stored in some manner, then recovered at a later time and converted back into a replica of the original sound—see Fig. 7.1.

Thus the process of recording and playback of music has three main elements: initial capture of sounds, usually using a microphone, then storage, for instance on tape, CD, or on a computer, and finally reconstruction of the sound from the stored signal using a loudspeaker, headphones, or earphones. At various stages along the chain we may also employ amplifiers that increase the strength of electrical signals—microphones, for instance, typically produce signals too small for storage, so the signals must be amplified first.

Closely related to sound recording are the processes of sound reinforcement and broadcasting. Sound reinforcement, the amplification of sound for live performance,

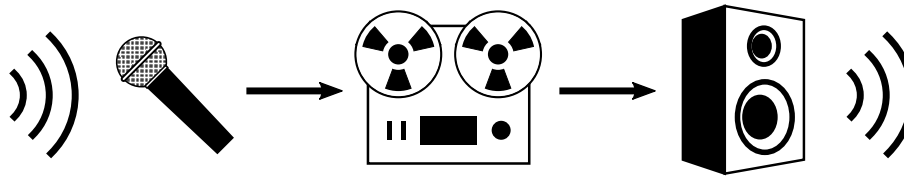


Figure 7.1: The recording process. Sound is turned into an electrical signal, usually with a microphone, then the electrical signal is stored in some manner, such as on magnetic tape or (more commonly) on a computer. At playback, the signal is recovered from storage and played through a loudspeaker, headphones, or earphones to recreate the original sound.

involves the same steps as recording except that the signal is never stored but goes straight to an amplifier and loudspeaker for immediate playback. Broadcasting, over radio or the Internet, involves an additional step of transmitting the signal to a remote receiver before the sound is reconstructed for the listener.

In this chapter we study the science behind the recording process, including the workings of microphones, loudspeakers, and various types of recording media, but sticking for the moment to traditional analog technologies. Digital recording, which works in a completely different manner, is discussed separately in the following chapter. We do not discuss amplifiers in detail in this chapter: their design and workings are largely the domain of electrical engineering and are beyond the scope of the book, though we will discuss the musical use of amplifiers, particularly in the context of the electric guitar, in Chapter 14.

First, however, in preparation for our discussion—and because it will also be useful in later chapters—we need to learn about circuits and the use of electrical signals to represent sound.

7.1 Electricity and sound

Modern music recording and amplification technology operates by turning sound into an equivalent electrical signal. We have already studied in some depth the science behind sound waves. To understand recording and amplification we need to understand electrical signals too.

Electricity consists of a flow of particles—electrons—each of which carries a certain amount of electric charge, the same on every electron. Electrons are too small to be visible to the eye, and the amount of charge they carry is also small, but a large enough number of them—and there are many trillions in a typical electrical cable or wire—can add up to a large current.

An electrical plug has two prongs and electricity flows into one and back out of the other, completing a circuit. Some plugs have three prongs, but the third (ground) prong is mainly a safety device—it doesn't normally play a role in the circuit.

Electricity flows easily through metals and a few other materials such as carbon, silicon, and water, but not through most other substances. This is convenient because it allows us to funnel electricity in any desired direction by channeling it along metal wires. The number of electrons that can fit in such a wire is large but not infinite and, because there is a limit on this number, electricity behaves in some ways like water flowing in a pipe: you can put electricity in one end of a wire only if you take a matching amount out of the other end. For this reason, electricity always flows in a loop or *circuit*, with an entrance and an exit. This is why a battery has two terminals, for instance: one terminal feeds electricity into the circuit and the other takes it out again. It's also why an electrical wall plug has two prongs.

Talking of batteries, every circuit needs a source of electricity. Electrons will not flow on their own but need to be pushed. The push might come from a battery, a generator, or a wall socket (which in turn gets electricity from a far-away genera-

tor). A power source such as a battery has an *electric potential*, also called a *voltage* and measured in volts, which is a measure of the potential for that source to drive electricity around a circuit. Electric potential is always measured between two different points, such as the two terminals of a battery, and a source with a high electric potential, like a 110 volt wall socket, can produce more electric flow than one with low potential, like a 1.5 volt battery. The electric potential does not measure how much electricity is actually flowing in a circuit. Rather, it is a measure of how much flow one could potentially get. Nonetheless, it plays an important role. If you want to know how big an electric shock you will get from an outlet, for instance, then it is the voltage you need to know about: you need to know whether the outlet has the *potential* to create a large flow of electricity when you touch it. If it doesn't then you will not get a shock.

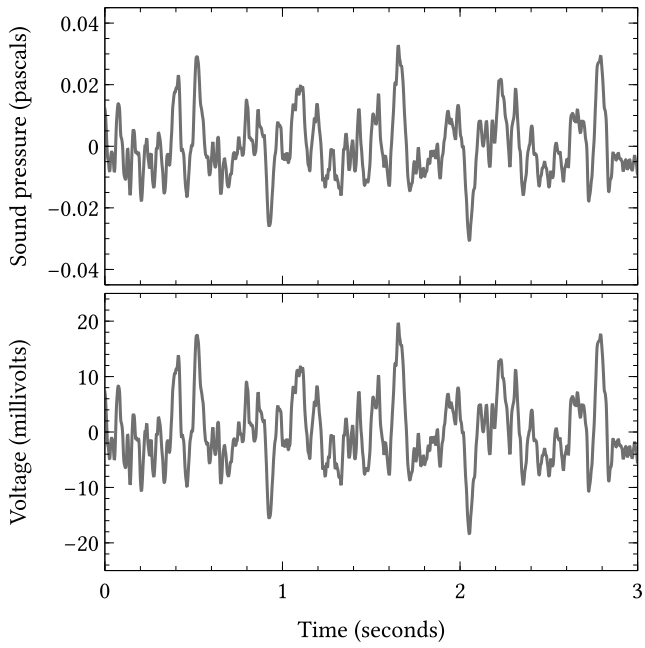


Figure 7.2: Representation of a sound waveform as an electrical signal. At the most fundamental level, all sound technology represents sound waveforms as varying voltages of size from a few millivolts up to a few volts. The shape of the sound waveform (top) is exactly mirrored in the variation of the voltage (bottom).

The actual amount of electricity flowing around a circuit is called the *current*. Current is measured at a single point, which can be anywhere in the circuit, and answers the question "How much electricity is flowing past this point per second?" In principle, one could measure current in terms of number of electrons per second, but following long-established tradition (going back to well before we even knew what electrons were) current is actually measured in *amps*. Technically an amp is equal to a flow of 6.2×10^{18} electrons per second, or six billion billion, though this will not be important for our purposes. An amp is merely a convenient unit of current that everyone agrees on.

7.1.1 Sound as an electrical waveform

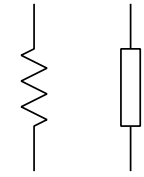
Recording technology captures sound waveforms as electrical signals by varying the *voltage* to represent the shape of the waveform. Take a look at Fig. 7.2. The top graph shows the sound pressure for a particular sound. The bottom one shows the corresponding voltage varying over time with the exact same pattern. When one goes up or down, so does the other. This is how sound is converted into an electrical signal. Given the right equipment we can also turn it back into sound again and this is the

principle behind all recording, amplification, and broadcast of sound and music.

A resistor

7.2 Electrical circuits

Electrical circuits of the kind used in music recording and amplification typically consist of a set of *components*, arranged around one or more loops, along with a power source to drive the electricity, such as a battery. The three main types of components found in circuits are resistors, capacitors, and transistors. Let us look at what each of these does, starting with resistors.



Two alternative symbols use to represent resistors in circuit diagrams.

7.2.1 Resistors

A *resistor* is an electrical component that, as its name suggests, impedes or resists the flow of electricity, reducing the amount that flows around a circuit. At first glance this might seem like a strange thing to do, but in most cases we do not want large electric currents flowing around our circuits. Normally it's more convenient, not to mention safer, to work with small ones, and resistors play a central role in almost all circuits by allowing us to reduce and control the size of the flow.

Physically, a resistor is a small cylinder a few millimeters long with two wires coming out of its two ends. When we draw pictures of electrical circuits, however, we conventionally represent resistors and other components by stylized symbols. The symbol for a resistor is either a zigzag line or a rectangular box, with straight lines coming out the ends to represent the two connecting wires. Both symbols are widely used. Here we will use the zigzag line.

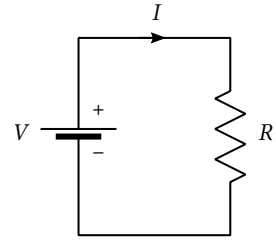


Figure 7.3: A simple circuit. This circuit consists of a battery with voltage V on the left and a single resistor with resistance R on the right. The battery produces a current I that flows around the circuit in the direction shown and through the resistor.

Figure 7.3 shows perhaps the simplest possible circuit we could have, which consists of just two components, a battery and a resistor. The battery is represented by the symbol on the left, consisting of two bars, one long and one short, and the resistor is on the right.

The amount by which a resistor resists the flow of electricity is called its *resistance* and is measured in ohms, denoted by the Greek symbol Ω ("omega"). Thus we might say a resistor had resistance 250 ohms or 250 Ω. The higher the resistance of a resistor, the less current will flow through it, with the current obeying *Ohm's law of resistance*.[1] Referring to Fig. 7.3, the voltage across the resistor, from one side to the other, is in this case just equal to the voltage V of the battery, and if the resistor has resistance R ohms then Ohm's law says that the current in amps flowing around the

[1]Not to be confused with Ohm's law of acoustics, discussed in Section 4.5, which refers to the ear's inability to hear changes in the phase of a sound.

circuit and through the resistor is

$$I = \frac{V}{R}, \tag{7.1}$$

with the current flowing from the positive side of the battery around the circuit to the negative side, as shown by the arrow in the figure. This law applies to all resistors: the current in a resistor is always equal to the voltage across the resistor divided by the resistance.

Suppose for example that we have a 1.5 volt battery and a resistor of 250 Ω. Then the current flowing around the circuit and through the resistor will be

$$I = \frac{1.5}{250} = 0.006 \text{ amps} = 6 \text{ milliamps}, \tag{7.2}$$

which is typical of the small currents used in audio circuits. One amp is a very large current. Currents of a few milliamps are more normal.

Note that current is conventionally denoted by the letter I and not, for instance, by C. The use of an I has its origin in the French word for current, *intensité*.

The equation for Ohm's law, Eq. (7.1), tells us that the current increases with the voltage V—more volts mean more current—but decreases as the resistance gets larger. Thus we can use the resistor to regulate the current flowing in our circuit. Resistors are manufactured with resistance values everywhere from one ohm to billions of ohms, allowing us wide latitude in our choice of current.

7.2.2 The potential divider or volume control

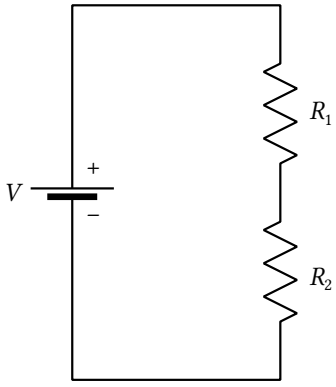


Figure 7.4: A circuit with two resistors. This circuit consists of a battery with voltage V and two resistors in series with resistances R_1 and R_2.

Now consider the circuit shown in Fig. 7.4. This circuit is more complicated in that it has two resistors with resistances R_1 and R_2 and the current from the battery has to pass through both resistors in order to complete the circuit. We say the resistors are wired *in series*. Resistors in series obey a simple rule: each one resists the flow according its own resistance and the combined effect is that of a single resistor with resistance $R_1 + R_2$. That is, we simply add the resistance values together.

This allows us to calculate the current flowing around this circuit using Ohm's law, Eq. (7.1). We set $R = R_1 + R_2$ and get

$$I = \frac{V}{R_1 + R_2}. \tag{7.3}$$

Suppose, for instance, that the battery again has voltage 1.5 volts and the two resistors have the same resistance $R_1 = R_2 = 250\,\Omega$. Then the current around the circuit is

$$I = \frac{1.5}{250 + 250} = 0.003 \text{ amps} = 3 \text{ milliamps}. \tag{7.4}$$

But now let us ask a different question: what is the voltage V_1 across the resistor R_1 if we measure the electrical potential between its two terminals? (Recall

that voltages are always measured between two points.) We can answer this question by again using Ohm's law, which links the current flowing in the resistor to the voltage across it. The same current I flows through both resistors and for resistor R_1 the current is related to the voltage V_1 according to $I = V_1/R_1$. Rearranging, we find that

$$V_1 = IR_1 = 0.003 \times 250 = 0.75 \text{ volts}. \tag{7.5}$$

We can do the same calculation for the voltage across the second resistor and we find that

$$V_2 = IR_2 = 0.003 \times 250 = 0.75 \text{ volts}. \tag{7.6}$$

Thus each of the resistors has a smaller voltage across it in this case, both of size 0.75 volts, which is half the size of the original 1.5 volts from the battery. This means that the two voltages on the resistors add up to the original battery voltage: 0.75 + 0.75 = 1.5 volts. This is not a coincidence: it is always true in any circuit, no matter how it is wired or what components it contains. The voltages always add up in this way. This will be a useful rule of thumb. We will see that we can use it to quickly calculate unknown voltages.

If we change the voltage on the battery, the same principles still apply. Whatever voltage comes out of the battery in Fig. 7.4, we will get a voltage across each of the resistors that is a half the size. More generally, for any choice of the resistances R_1 and R_2, the current I is given by Eq. (7.3) and the voltage across R_2 is

$$V_2 = IR_2 = \frac{R_2}{R_1 + R_2} V. \tag{7.7}$$

Thus this circuit acts as a *potential divider*: it takes whatever voltage you start with and cuts it down by a factor of $R_2/(R_1 + R_2)$. By choosing the values of the two resistors we can arrange this factor to be any number we like between zero and one. This is useful for instance if you need a smaller voltage than the one the battery gives you. If you have a 1.5 volt battery but you want 0.75 volts then you can achieve this by using a potential divider.

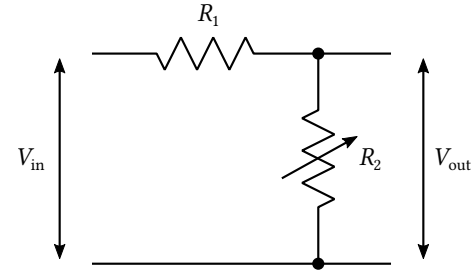


Figure 7.5: A volume control. The potential divider circuit of Fig. 7.4 can be rearranged to act as a volume control. It takes an input waveform in the form of a time varying voltage V_{in} and cuts it down by a factor of $R_2/(R_1 + R_2)$, which can be varied by changing the resistance R_2.

Taking things a step further, suppose the voltage on the left of the circuit is not a steady voltage from a battery at all, but a time-varying voltage representing a musical waveform, as discussed in Section 7.1. The divider will take that voltage and cut it down by precisely the fraction $R_2/(R_1 + R_2)$. The shape of the waveform will stay the same but everything about it will be reduced in size by this same factor. This smaller waveform will then produce quieter music when we turn the signal back into sound again, so the circuit is acting as a volume control.

When used in this way, the potential divider circuit is usually drawn slightly differently, as shown in Fig. 7.5. It's still the same circuit, just rearranged. The input waveform arrives on the left as shown and the output waveform is produced on the right, with

$$V_{\text{out}} = \frac{R_2}{R_1 + R_2} V_{\text{in}}. \tag{7.8}$$

We have also made one other change to the circuit: the zigzag line with an arrow through it is the symbol for a *variable resistor*, a standard component that acts as a resistor but has a knob that you can turn to change the amount of resistance. By turning the knob we can vary R_2 and hence vary the amount the circuit cuts the signal down by. This allows us to vary the volume of the sound all the way from full volume—the entire original signal—when R_2 is very large, to zero when it is very small. This is precisely how a volume control operates.

EXAMPLE 7.1: VOLUME CONTROL

A volume control like the one in Fig. 7.5 has $R_1 = 1200\ \Omega$ and a variable resistor for R_2. If the input voltage is $V_{\text{in}} = 1$ volt and R_2 is set to $400\ \Omega$, what is the output voltage?

To answer this question we apply Eq. (7.8), which gives

$$V_{\text{out}} = \frac{R_2}{R_1 + R_2} V_{\text{in}} = \frac{400}{1200 + 400} \times 1 = 0.25 \text{ volts}. \tag{7.9}$$

Now what value would we have to set R_2 to if we wanted the output voltage to be 0.6 volts? To answer this question, we rearrange Eq. (7.8) to give R_2 thus:

$$R_2 = \frac{V_{\text{out}}}{V_{\text{in}} - V_{\text{out}}} R_1. \tag{7.10}$$

Plugging in the values we then have

$$R_2 = \frac{0.6}{1 - 0.6} \times 1200 = 1800\ \Omega. \tag{7.11}$$

7.2.3 CAPACITORS

A capacitor

A *capacitor* is an electronic component that stores a small amount of electrical charge. More correctly, it stores two charges, one positive and one negative, of equal size. (Electrical charge can be both positive and negative. Like savings and debt, positive and negative charges cancel out when they meet.) Physically a capacitor consists of two metal plates a short distance apart inside a casing, often with a fluid or other insulating medium in between. In practice you can't see the plates, just the casing, and there are two wires, one attached to each plate, that are used to connect the capacitor to a circuit.

When you connect a capacitor to a battery or other power source, a current flows, pushing electrons onto one plate of the capacitor and pulling an equal number off

the other. There is a limit to how much the plates will take, however, and when that limit is reached no more current will flow, so in practice what we see is an initial burst of current, then nothing after that.

The result is an electric charge on the two plates in the form of excess electrons on one and a corresponding deficit on the other. The charges on the plates are equal and opposite, so the net amount of charge on a capacitor is actually always zero, but the amount on each plate is not.

On the plate connected to the positive side of the battery there is a positive charge. As with current we could measure the charge by counting electrons, but by long-standing tradition we instead measure it in units of *coulombs* (denoted C). A coulomb is the charge of 6.2×10^{18} electrons. Let us denote by Q the charge in coulombs on the positive plate. The other plate, the one connected to the negative side of the battery, then has a negative charge of $-Q$ in coulombs.

The specific value of Q for a capacitor depends on the applied voltage V according to the capacitor law

$$Q = CV, \tag{7.12}$$

where C is the *capacitance*, a measure of the storage ability of the capacitor. Capacitance is measured in *farads* (denoted F). One farad is a large amount of capacitance: real capacitors come nowhere near this capacity. A typical value in practice is one microfarad, i.e., one millionth of a farad, denoted μF.

Now take a look at the circuit in Fig. 7.6, which contains a battery, a resistor, and a capacitor. The symbol for a capacitor is two parallel lines as shown—a simplified representation of the two plates. Suppose that initially the battery is not connected but then we connect it up to power the circuit. What will happen?

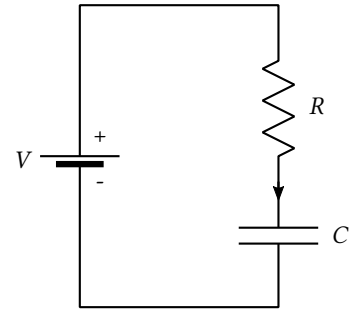


Figure 7.6: A circuit using a capacitor. This circuit consists of a battery with voltage V and a resistor R and capacitor C.

Before the battery is connected there will be no charge on the capacitor, so $Q = 0$. Rearranging Eq. (7.12), we find that $V = Q/C$, so the voltage across the capacitor will be zero as well, which makes sense since there is no power source to produce any voltage yet. Now we connect the battery. When we first connect the battery the resistor must have the entire voltage V across it because, as we have said, voltages always add up—the voltages across the resistor and the capacitor have to add up to V and the voltage across the capacitor is zero. But if there is voltage V across the resistor then we will have a current flowing through the resistor as given by Ohm's law $I = V/R$ (Eq. (7.1)).

As this current flows out of the resistor it flows into the capacitor and charges it up as described above, and as the capacitor charges up the voltage across it will start to grow: the equation $V = Q/C$ tells us that the more charge we have the larger the voltage. This means that the voltage across the resistor will *decrease* (because the voltages have to add up to V) and hence the current flowing

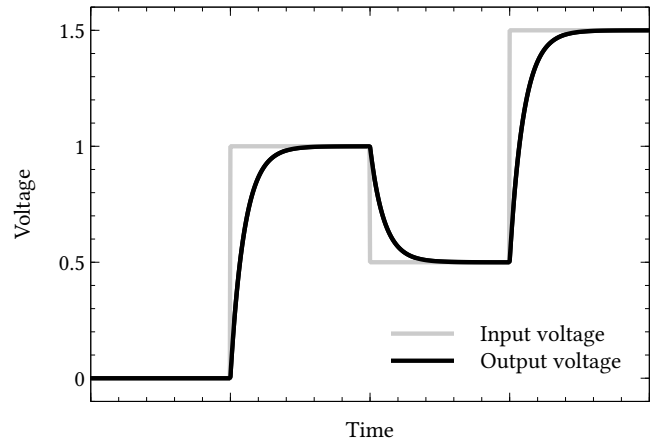


Figure 7.7: The effect of the circuit in Fig. 7.6. When the input voltage V changes to a new value, the output voltage—the voltage across the capacitor—follows, but with a time lag.

will also decrease.

Overall, therefore, we get an initial rapid flow of current, charging the capacitor, but over time the current will die off. Eventually, when the charge on the capacitor becomes large enough, the current will stop altogether. This happens when $Q = CV$, so that the entire voltage V is across the capacitor and the voltage across the resistor is zero, producing zero current.

Another way to say this is that the voltage across the capacitor copies the voltage of the battery, but with a lag. It starts at zero when there is no battery connected, then after the battery is connected it rises, until it reaches the battery voltage and then it stops there. Figure 7.7 shows what this looks like in graph form.

Now suppose the voltage of the battery changes. This does not normally happen with actual batteries, but the voltages used to represent sound signals vary a lot, so let's take a look. If the voltage changes, then again the voltage on the capacitor will copy that change but with a lag—see Fig. 7.7 again. If the voltage changes repeatedly, as shown in the figure, then the capacitor will copy every change, but with a lag.

The size of the lag is determined by how long it takes the capacitor to charge up. If the capacitor has a very large capacitance C it will take longer. Twice the capacitance means it takes twice as long, so the lag is proportional to C. At the same time it also depends on the resistance R, since more resistance means less current and hence slower charging. Twice the resistance means twice as slow, so the lag is also proportional to R. In Section 7.2.5 we show that in fact the time lag, also called the *time constant* of the circuit, is equal to RC. So with a 1 μF capacitor, for instance,

and a 2000 Ω resistor (also called 2 kilohms or 2 kΩ) the time constant would be

$$RC = 2000 \times 10^{-6} = 0.002 \text{ seconds.} \tag{7.13}$$

7.2.4 Filters

The circuit of Fig. 7.6 turns out to be musically useful. To see why, let's start by rearranging it into the form shown in Fig. 7.8. This is just the same circuit as before, but reorganized to accept an input voltage V_{in} and produce an output voltage V_{out}, similar to the way we drew the potential divider circuit in Fig. 7.5. Typically the input voltage would be a signal representing a sound waveform. And what would the output voltage look like? Figure 7.9 shows two examples. In both, the input voltage is a sine wave, but with different frequencies. In the top panel the frequency is a slow 20 Hz, meaning we have one cycle every 0.05 seconds, at the very low end of the musical frequency range. Meanwhile, we have made the time constant of the circuit—the size of the lag between output and input—equal to $RC = 0.002$ seconds as above. Thus, in this case, the lag is much shorter than the period of the wave and hence has very little effect. The output voltage lags just a short distance behind the input and ends up following the input voltage closely, so the difference between the two is barely visible in the figure.

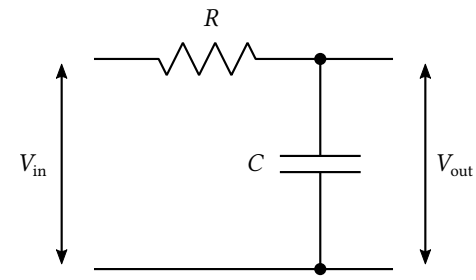


Figure 7.8: A low-pass filter. The circuit of Fig. 7.6 reorganized as a low-pass filter.

In the lower panel of the figure the input sine wave has a much higher frequency of 2000 Hz, near the highest of musical notes, while the time constant of the circuit is still 0.002 seconds. Now we have one cycle of the wave every 0.0005 seconds, much less than the time constant. This means that the output voltage does not have nearly enough time to catch up with the input. It needs around 0.002 seconds to catch up, but the input signal changes long before this as it goes through its cycle. The result is that the output never manages to catch up with the input, so the output signal is much smaller than the input.

Overall then, the effect of this circuit is that when we feed in a low-frequency signal it comes out almost unchanged, but when we feed in a high-frequency one very little gets through to the output. We call such a circuit a filter. Specifically, this is a *low-pass filter*, a circuit that lets low-frequency signals though but cuts out high-frequency ones.

Now recall that, as discussed in Section 4.3, any sound can be represented as a combination of sine waves. Each sine wave in a sound will be affected by the filter according to its own particular frequency, so for a complex waveform made of many sine waves the filter will cut down on the high-frequency ones and leave the low-

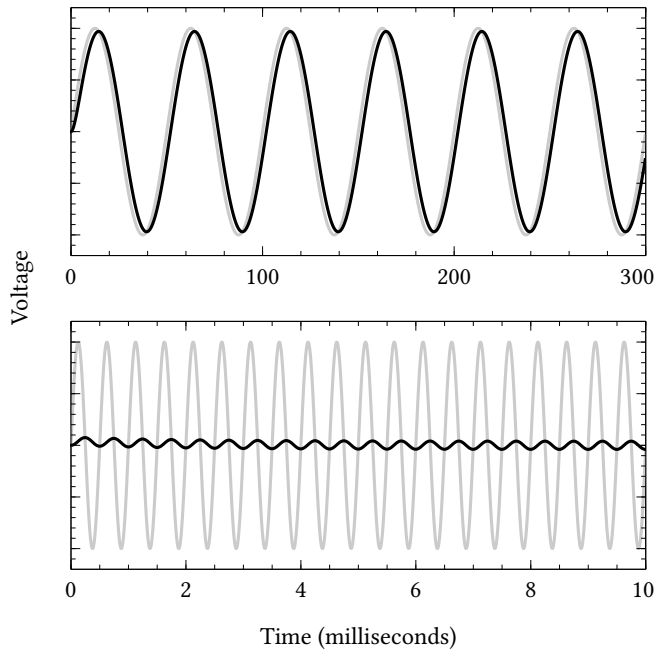


Figure 7.9: The effect of the low-pass filter. Top: when a low-frequency sine-wave signal is fed into the circuit of Fig. 7.8 it emerges from the output essentially unchanged—there is a slight lag between the two, but the effect is small. Bottom: when a high-frequency signal is fed in only a small signal comes from the output. Note that the horizontal scale is different in the two panels.

frequency ones unchanged. In musical terms, this has the effect of cutting down the amount of treble while leaving the bass as it is. This is extremely useful for many purposes. Filters like this are used for instance as tone controls on amplifiers, for sound shaping in electronic synthesizers, and for many things in recording studios.

With a little mathematics we can calculate the exact effect of our low-pass filter on an incoming electrical signal. The details are given in Section 7.2.5 but the crucial result is that for a sine-wave input the intensity of the corresponding sound obeys the equation

$$I_{\text{out}} = \frac{I_{\text{in}}}{1 + (2\pi RCf)^2}. \tag{7.14}$$

Whatever the intensity I_{in} of the input signal, the output signal has the same intensity but divided by $1 + (2\pi RCf)^2$. Figure 7.10 shows what this looks like as the frequency varies. Look first at the middle curve in the figure. At low frequencies I_{out} is equal to I_{in}—the filter has no effect and whatever goes in comes out unchanged. But as the

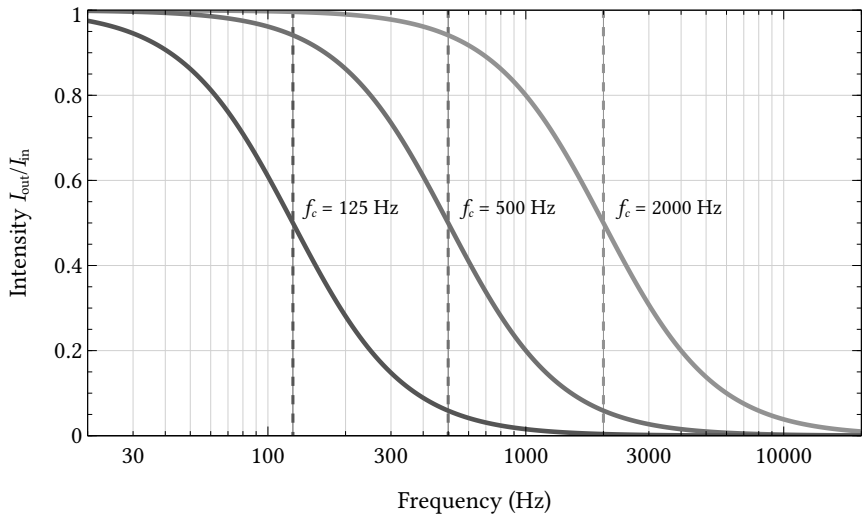


Figure 7.10: Frequency response of a low-pass filter. The intensity of the output signal from the low-pass filter of Fig. 7.8 as a function of frequency for three different values of the cutoff frequency: $f_c = 125\,\text{Hz}$, 500 Hz, and 2000 Hz.

frequency gets higher the intensity starts to fall off and for the highest frequencies it is virtually zero.

We can work out the frequency where the intensity falls off by looking at Eq. (7.14). If $f = 0$ we see that $I_{\text{out}} = I_{\text{in}}$, so the filter has no effect, while if f is very large then I_{out} will be almost zero. In between these two lies a point where $I_{\text{out}} = \frac{1}{2}I_{\text{in}}$, so that the filter cuts the intensity down by exactly a half. This half-way point is called the *cutoff frequency* of the filter, denoted f_c. It's straightforward to see that the cutoff frequency corresponds to the point where $(2\pi RCf)^2 = 1$ and, rearranging for f, we find that

$$f_c = \frac{1}{2\pi RC}. \tag{7.15}$$

Roughly speaking, sound below this frequency gets through the filter and sound above it gets blocked.

We can vary the cutoff frequency of the filter by varying either R or C (or both). Figure 7.10 shows curves for three different choices, $f_c = 125\,\text{Hz}$, 500 Hz, and 2000 Hz, as indicated by the vertical dashed lines. As we can see, the curve simply gets shifted from side-to-side as the cutoff frequency varies.

Another way to look at the low-pass filter is in terms of the number of decibels

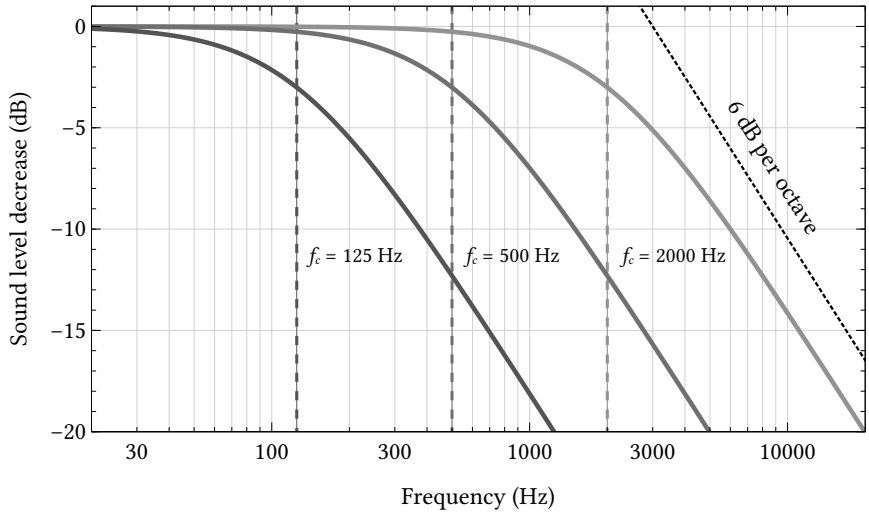


Figure 7.11: Sound level decrease upon applying the low-pass filter. The amount in decibels by which the low-pass filter of Fig. 7.8 decreases the sound level of a signal, as a function of frequency for three different values of the cutoff frequency: $f_c = 125\,\text{Hz}$, 500 Hz, and 2000 Hz.

it cuts out. Combining Eqs. (7.14) and (7.15) we can write

$$\frac{I_{\text{out}}}{I_{\text{in}}} = \frac{1}{1 + f^2/f_c^2}\,, \tag{7.16}$$

and substituting into Eq. (3.16) on page 60, we find the change in the decibel level of the sound to be

$$L_{\text{out}} - L_{\text{in}} = -10 \log\left(1 + \frac{f^2}{f_c^2}\right). \tag{7.17}$$

Figure 7.11 shows a plot of this formula for the same values of f_c as in Fig. 7.10, with the cutoff frequency corresponding roughly to the point where the filter transitions from allowing the signal to pass unchanged (zero decibels reduction—the flat part of the curve on the left) to cutting it off (the falling curve on the right).

Above the cutoff frequency, where f/f_c is much greater than one, we can ignore the 1 in the denominator of Eq. (7.16) and to a good approximation

$$\frac{I_{\text{out}}}{I_{\text{in}}} = \frac{f_c^2}{f^2}\,, \tag{7.18}$$

so the output intensity well above f_c is inversely proportional to f^2. Thus if the frequency doubles the intensity goes down by a factor of four, or equivalently by 6 dB.

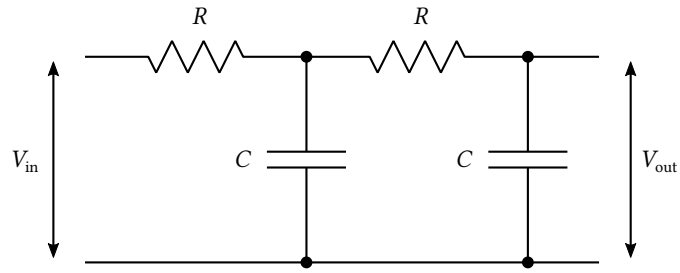


Figure 7.12: A 12 dB-per-octave low-pass filter. This filter consists of two low-pass filters, one after another, each cutting out 6 dB per octave, for a total of 12 dB per octave.

In other words, the sound level falls by 6 dB per octave. This behavior is shown as the diagonal dashed line on the right of Fig. 7.11 and, as we can see, the curves do indeed have roughly this slope.

There is a wide range of uses for low-pass filters, with all kinds of values of the cutoff frequency. Filters with very high cutoff around $f_c = 20\,000$ Hz are used as anti-aliasing filters and reconstruction filters to remove unwanted ultrasonic frequencies from music. Filters with intermediate values of f_c are used as tone controls on amplifiers and electric guitars, for equalization or "EQ" in studio recording and mixing, as sound shaping filters in synthesizers, and as crossover filters in loudspeaker systems. Filters with an ultra-low cutoff of 10 Hz or less are used as compensation filters for tape recorders, record players, or dynamic microphones, that otherwise produce too much high-frequency sound.

Sometimes we may want a filter that cuts out more of the high frequencies than the 6 dB-per-octave filter described here. In that case, we can string two or more filters together one after another, as shown in Fig. 7.12. If we have two filters and each one cuts out 6 dB per octave, then together they will cut 12 dB per octave. Three filters will cut 18 dB per octave and four will cut 24 dB per octave. You can find each of these filter types in use in certain situations. For instance, the distinctive sounds of different brands of electronic synthesizers are in part due to the different filters they use. Moog brand synthesizers have long made use of 24 dB-per-octave filters, but other successful brands such as Oberheim and Sequential Circuits use 12 dB-per-octave filters. This gives the former a warmer sound and the latter a brighter sound. (We study the workings of synthesizers and their use of filters in Chapter 15.)

Another variant of our filter circuit is shown in Fig. 7.13. The only difference between this and the low-pass filter is that the resistor and the capacitor are the opposite way around, but it turns out that this makes all the difference in the world. When we feed a sine-wave signal into the input of this circuit, the voltage across the capacitor decreases as frequency goes up just as before. But this means that the

voltage across the resistor—which plays the role of output signal in this circuit—must go down, since the two voltages always to have to add up to V_{in}. Thus this circuit is a *high-pass filter*. It produces a small signal at low frequencies but a large signal at high frequencies.

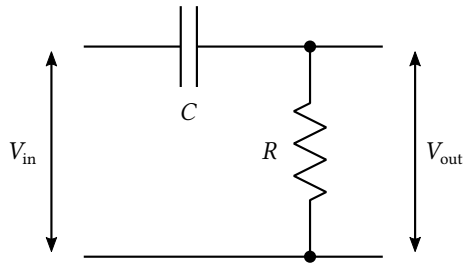


Figure 7.13: A high-pass filter. This circuit acts as a high-pass filter, cutting out 6 dB per octave at low frequencies but leaving high frequencies unchanged.

In Section 7.2.5 it is shown that the intensity of the output from the high-pass filter circuit is

$$I_{\text{out}} = \frac{(2\pi RCf)^2}{1 + (2\pi RCf)^2} I_{\text{in}} = \frac{I_{\text{in}}}{1 + f_c^2/f^2}, \tag{7.19}$$

where the cutoff frequency f_c is again given by

$$f_c = \frac{1}{2\pi RC}. \tag{7.20}$$

Figure 7.14 shows a plot of this output intensity against frequency—the equivalent for the high-pass filter of Fig. 7.10—and the high-pass behavior is clearly visible. At low frequencies the filter cuts 6 dB per octave, then it flattens out at frequencies above f_c. If a stronger effect is needed one can string two or more high-pass filters together to get 12 dB per octave or more.

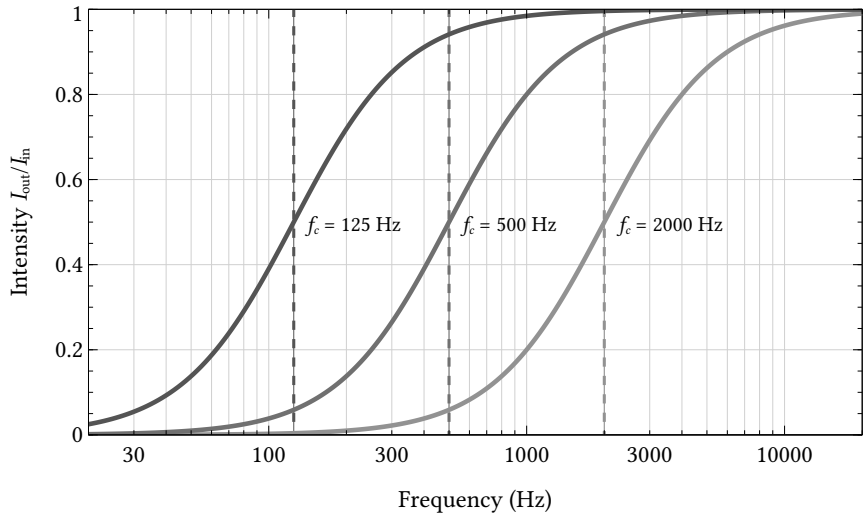


Figure 7.14: Frequency response of a high-pass filter. The intensity of the output signal from the high-pass filter of Fig. 7.13 as a function of frequency for cutoff frequencies f_c = 125 Hz, 500 Hz, and 2000 Hz.

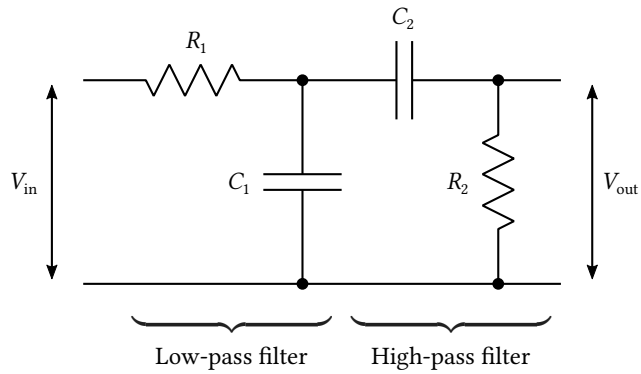


Figure 7.15: A band-pass filter. A band-pass filter consists of a low-pass filter and a high-pass filter, one after another.

High-pass filters are used as bass tone controls in amplifiers, for studio equalization, in crossover circuits for loudspeakers, and, more rarely, for sound shaping in synthesizers. A more specialized use of a high-pass filter, which will be important to us occasionally, is for filtering out constant voltages. If we reduce the frequency of an electrical signal all the way to zero the signal no longer changes and we have a constant voltage, so a constant voltage can be thought of as a signal with frequency zero. Sometimes we want a circuit that blocks constant signals while letting varying ones pass, and we can do this with a high-pass filter. Putting $f = 0$ in Eq. (7.19) we see that a high-pass filter will produce zero output from a constant input signal, no matter what the cutoff frequency f_c. So a high-pass filter with a very low cutoff—say 1 Hz—will block a constant signal but leave sound waveforms (with frequencies of 20 Hz and more) untouched. This trick is used, for example, in the condenser microphone—see Section 7.3.5.

A third type of filter is the *band-pass filter*, which consists of a low-pass filter and a high-pass filter, one after another, as shown in Fig. 7.15. The low-pass filter cuts out high frequencies and the high-pass filter cuts out low frequencies, leaving only the frequencies in the middle. By choosing the cutoffs of the low- and high-pass stages, f_{low} and f_{high}, one can tune the *bandwidth* of filter—the range of frequencies it allows to pass. Figure 7.16 shows the frequency response of the filter in Fig. 7.15 when the cutoffs are set at $f_{\text{low}} = 125\,\text{Hz}$ and $f_{\text{high}} = 2000\,\text{Hz}$. Band-pass filters find use in graphic equalizers and other studio tools for shaping and tuning recorded sound, in crossover circuits for loudspeaker systems, and for creating narrowband noise as described in Section 4.4.1.

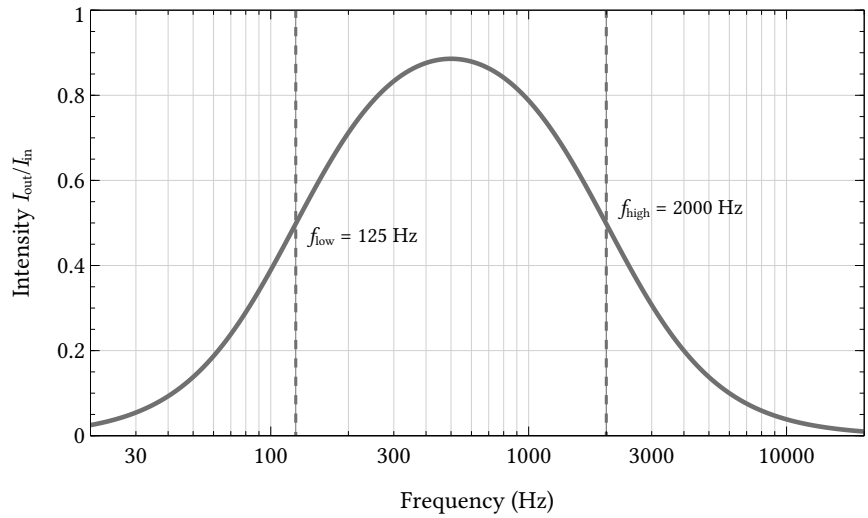


Figure 7.16: Frequency response of a band-pass filter. The intensity of the output signal from the band-pass filter of Fig. 7.15 as a function of frequency for cutoff frequencies $f_{\text{low}} = 125\,\text{Hz}$ and $f_{\text{high}} = 2000\,\text{Hz}$.

Example 7.2: A low-pass filter

A 6 dB-per-octave low-pass filter like the one in Fig. 7.8 has $R = 5100\,\Omega$ and $C = 0.1\,\mu\text{F}$. What is the cutoff frequency, and how many decibels does the filter cut from a signal at 1000 Hz? To answer this question, we first use Eq. (7.15) to calculate the cutoff frequency:

$$f_c = \frac{1}{2\pi RC} = \frac{1}{2\pi \times 5100 \times 10^{-7}} = 312\,\text{Hz}. \tag{7.21}$$

Now the change in decibels for a tone with frequency $f = 1000\,\text{Hz}$ is given by Eq. (7.17) to be

$$L_{\text{out}} - L_{\text{in}} = -10\log\left(1 + \frac{f^2}{f_c^2}\right) = -10\log\left(1 + \frac{1000^2}{312^2}\right) = -10.5\,\text{dB}, \tag{7.22}$$

so the sound level will drop by just a little over 10 dB.

Advanced material

7.2.5 Frequency response of filters

We can derive the frequency response of low-pass, high-pass, and band-pass filters from the equations governing circuits of resistors and capacitors. Take the case of the low-pass filter of Fig. 7.8. Suppose the current flowing around the circuit at a certain moment is I, so that the voltage across the resistor is IR, following Ohm's law, Eq. (7.1).

The voltages across the resistor and capacitor have to add up to the input voltage V_{in}, so

$$IR + V_{\text{out}} = V_{\text{in}}. \tag{7.23}$$

At the same time the charge Q on the capacitor obeys the capacitor equation, Eq. (7.12), so $Q = CV_{\text{out}}$. The rate at which charge flows onto the capacitor is equal to the current I (since current is by definition the rate at which charge flows) and hence

$$I = \frac{\mathrm{d}Q}{\mathrm{d}t} = C\frac{\mathrm{d}V_{\text{out}}}{\mathrm{d}t}. \tag{7.24}$$

Combining Eqs. (7.23) and (7.24) we then find that the output voltage V_{out} obeys the differential equation

$$RC\frac{\mathrm{d}V_{\text{out}}}{\mathrm{d}t} + V_{\text{out}} = V_{\text{in}}. \tag{7.25}$$

Now suppose the input to the circuit is a sine wave with frequency f, which we represent in conventional manner as the real part of the complex number $V_{\text{in}} = A_{\text{in}}\mathrm{e}^{\mathrm{i}2\pi ft}$, where A_{in} is the amplitude of the signal. Then the (steady-state) output of the circuit is another sine wave $V_{\text{out}} = A_{\text{out}}\mathrm{e}^{\mathrm{i}2\pi ft}$ and, substituting into Eq. (7.25) and cancelling some factors, we find that $\mathrm{i}2\pi RCfA_{\text{out}} + A_{\text{out}} = A_{\text{in}}$, or equivalently

$$A_{\text{out}} = \frac{A_{\text{in}}}{1 + \mathrm{i}2\pi RCf}. \tag{7.26}$$

In general this is a complex number. The actual real amplitude of the output signal is given by the magnitude

$$\bigl|A_{\text{out}}\bigr| = \frac{\bigl|A_{\text{in}}\bigr|}{\sqrt{1 + (2\pi RCf)^2}}. \tag{7.27}$$

When the electrical signal is transformed back into sound, it will produce an intensity proportional to the square of the amplitude (see Section 3.1.3), so $I = k|A|^2$ for some constant k. This means that

$$I_{\text{out}} = k\bigl|A_{\text{out}}\bigr|^2 = \frac{k\bigl|A_{\text{in}}\bigr|^2}{1 + (2\pi RCf)^2} = \frac{I_{\text{in}}}{1 + (2\pi RCf)^2}, \tag{7.28}$$

as stated in Eq. (7.14).

For the high-pass filter of Fig. 7.13 the analysis is similar. The voltages across capacitor and resistor must still add up to V_{in} and Eq. (7.12) tells us that the voltage across the capacitor is equal to Q/C, so

$$\frac{Q}{C} + V_{\text{out}} = V_{\text{in}}. \tag{7.29}$$

Differentiating with respect to time and recalling that $\mathrm{d}Q/\mathrm{d}t = I$ we then get

$$\frac{I}{C} + \frac{\mathrm{d}V_{\text{out}}}{\mathrm{d}t} = \frac{\mathrm{d}V_{\text{in}}}{\mathrm{d}t}. \tag{7.30}$$

At the same time, Ohm's law tells us that $I = V_{\text{out}}/R$, so we end up with the differential equation

$$RC\frac{\mathrm{d}V_{\text{out}}}{\mathrm{d}t} + V_{\text{out}} = RC\frac{\mathrm{d}V_{\text{in}}}{\mathrm{d}t}. \tag{7.31}$$

Again taking a sine-wave input in the form $V_{\text{in}} = A_{\text{in}}\mathrm{e}^{\mathrm{i}2\pi ft}$ and assuming a corresponding sine-wave output $V_{\text{out}} = A_{\text{out}}\mathrm{e}^{\mathrm{i}2\pi ft}$, we get

$$\mathrm{i}2\pi RCfA_{\text{out}} + A_{\text{out}} = \mathrm{i}2\pi RCfA_{\text{in}}, \tag{7.32}$$

or

$$A_{\text{out}} = \frac{\mathrm{i}2\pi RCf}{1 + \mathrm{i}2\pi RCf}A_{\text{in}}. \tag{7.33}$$

The real amplitude is then

$$\bigl|A_{\text{out}}\bigr| = \frac{2\pi RCf}{\sqrt{1 + (2\pi RCf)^2}}\bigl|A_{\text{in}}\bigr|, \tag{7.34}$$

and the intensity of the resulting sound obeys

$$I_{\text{out}} = \frac{(2\pi RCf)^2}{1 + (2\pi RCf)^2}\,I_{\text{in}}, \tag{7.35}$$

as stated previously in Eq. (7.19).

Other filters can be built from combinations of the basic low- and high-pass filters. For instance, if we string two low-pass filters together one after another as in Fig. 7.12, the output of the first becomes the input to the second. Each filter then multiplies the intensity by the factor in Eq. (7.28) and we get

$$I_{\text{out}} = \frac{I_{\text{in}}}{[1 + (2\pi RCf)^2]^2}, \tag{7.36}$$

which gives us a 12 dB-per-octave low-pass filter by contrast with the 6 dB per octave we get out of a single-stage filter. If we string a low-pass and a high-pass filter together as in Fig. 7.15 we get

$$I_{\text{out}} = \left(\frac{1}{1 + (2\pi R_1C_1f)^2}\right)\left(\frac{(2\pi R_2C_2f)^2}{1 + (2\pi R_2C_2f)^2}\right)I_{\text{in}}, \tag{7.37}$$

which gives the band-pass frequency response shown in Fig. 7.16.

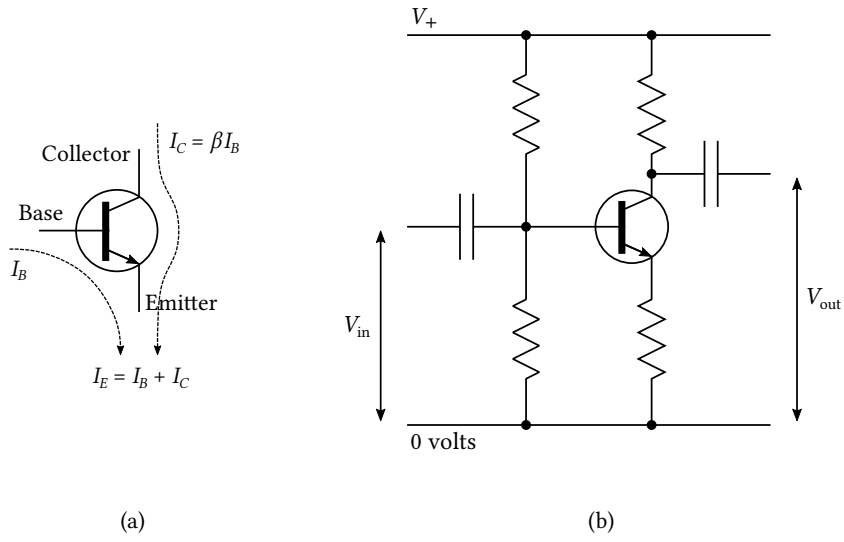
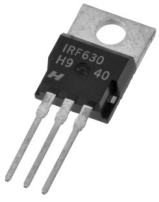


Figure 7.17: Transistor circuits. (a) When a small current I_B flows through the base of a transistor, it causes a much larger current $I_C = \beta I_B$ to flow through the collector. (b) A typical circuit diagram for a single-stage transistor amplifier. In addition to the transistor itself, shown in the center, the two resistors on the left act as a potential divider to center the voltage at the base of the transistor, while the two on the right act to convert currents into voltages via Ohm's law. The two capacitors act as high-pass filters to remove constant voltages from the input and output signals. The top line or "rail" is attached to a battery or other power supply with voltage V_+, which provides power to the circuit.

7.2.6 Transistors



A transistor

The third basic component of electronic circuits is the *transistor*. Invented at Bell Labs in the 1940s, the bipolar junction transistor is a device with three connections—not just two like the resistor and capacitor—whose function is to amplify electrical signals, taking small signals and making them bigger. This is a crucial step in sound recording and live audio and the invention of the transistor revolutionized how music is recorded and performed.[2]

A transistor is represented in electrical circuits by the circular symbol shown

[2]Before the invention of the transistor, electronic devices made use of a different technology, the *vacuum tube* or *valve*, which can also be used as an amplifier but is significantly inferior to the transistor in many ways and substantially limits the practical amount of power that amplifiers can generate. Vacuum tubes have been mostly replaced by transistors in recent decades but do still have some specialized uses, in part because of their imperfections. For instance, we discuss their use in distorting guitar amplifiers in Section 14.2.

in Fig. 7.17a. The three connections to the device each play different roles and are called the *base*, *collector*, and *emitter*. In normal use, one applies a voltage between the collector and emitter, but initially no current flows because the transistor has a high resistance. If, however, one sends a small current into the base and out through the emitter, then the resistance of the device changes and current starts to flow from collector to emitter. The crucial observation is that the current I_C flowing through the collector is much larger than the current I_B flowing through the base, so that the transistor acts as an amplifier. Specifically,

$$I_C = \beta I_B, \tag{7.38}$$

where the quantity β is known as the *gain* of the transistor. A typical value of the gain is $\beta = 100$ or so, so the effect is a large one: you can feed in a current of one milliamp and get out a current of 100 milliamps. By combining several transistors, one after another, one can create circuits with a huge overall gain, a factor of thousands or even millions, that can take the tiniest signals, such as those from microphones or electric guitars, and make them large enough to fill entire arenas.

Transistors are not normally used on their own, but rather in combination with a selection of resistors and capacitors to make a complete circuit. Figure 7.17b shows a typical circuit for a one-transistor amplifier. We will not discuss the design of transistor circuits in detail here—it is a complex topic, the subject of entire books just on its own—but we will discuss the use of musical amplifiers of various kinds in Chapter 14.

7.2.7 Electromagnets and induction

In addition to resistors, capacitors, and transistors, recording technology uses a range of more specialized components specifically for processing sound, such as microphones and loudspeakers, which make use of electromagnetic effects for their operation. An *electromagnet* is a magnet powered by electricity. If one takes a loop of wire, as shown in Fig. 7.18, and passes a current around it (for instance by connecting it to a battery), the current creates a magnetic field through the loop.

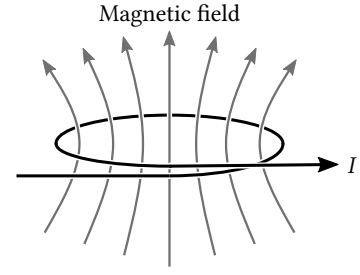


Figure 7.18: An electric current in a loop of wire produces a magnetic field.

The magnetic field from a single loop of wire, however, is very small—barely large enough to be detectable and not of much use. But if you put two loops together you get twice the field, three loops gives three times, and so forth. By putting many loops together to form a coil one can create a powerful electromagnet with a field strong enough to attract metal and magnetize other objects. The field can be made even stronger by placing a metal core inside the coil. The core becomes magnetized by the field and amplifies the effect of the electromagnet. Powerful electromagnets are used in hospital MRI machines to scan patients' insides, in electromagnetic door locks, and

to pick up scrap metal in junk yards. They also play a key role in music technology as a central component of tape recorders, computer hard drives, and loudspeakers.

A practical electromagnetic might consist of hundreds or thousands of loops of wire around a metal core, creating the coil called a *solenoid*—see Fig. 7.19. The strength B of the magnetic field produced when current I flows around a solenoid is measured in units of teslas and given by the formula

The tesla unit (like the automobile of the same name) is named after Serbian engineer Nikola Tesla, who was responsible for numerous inventions in the early days of the commercial electricity industry.

$$B = \frac{\mu N I}{L}. \tag{7.39}$$

Here L is the length of the solenoid from end to end, N is the number of turns in the coil (how many times it loops around the core), and μ is a quantity called the *magnetic permeability*, which measures how much the core amplifies the field. If the core is made of a material like iron, which has a large permeability, then the field is greatly amplified and we will get a strong electromagnet. The core can be extended past the ends of the coil as shown in Fig. 7.20 or bent around to carry the magnetic field where it is needed, a useful trick in technological applications where space is cramped. (When we do this the length L appearing in Eq. (7.39) is the length of the coil, not the length of the core.)

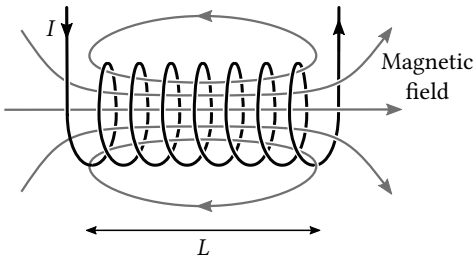


Figure 7.19: A solenoid. A coil with many turns produces a strong magnetic field, in proportion to the number of turns, when a current I passes through it.

Equation (7.39) tells us that the strength of the magnetic field produced by a solenoid is proportional to the current I. This is crucial in audio applications of electromagnets. In a loudspeaker, for instance, as we will see in Section 7.5.1, an electrical signal is sent through an electromagnet, producing a magnetic field that moves a diaphragm, which in turn pushes on the air and makes a sound. Because the strength of the electromagnet is proportional to the current in the electrical signal, the movement of the diaphragm is also proportional to the current, and hence the resulting sound mimics the waveform of the electrical signal—exactly what we need to turn electricity into sound.

Related to electromagnets is the phenomenon of *electromagnetic induction*, which is in some ways the reverse of what an electromagnet does. In an electromagnet a current passed through a coil generates a magnetic field; in electromagnetic induction a magnetic field passes through an otherwise inert coil which *induces* an electric voltage in the coil. The two processes are not quite analogous though, for two reasons. First, it is *current* that produces the field in an electromagnet but induction generates a *voltage*. Second, with an electromagnet the strength of the magnetic field depends on the size of the current, but with induction the voltage depends not directly on the strength of the field but on how fast the field is changing. A rapidly changing magnetic field produces a large voltage, a slowly

changing field produces a small voltage, and a stationary magnetic field, one that is not changing, will produce no voltage at all, no matter how strong it is. This turns out to be a crucial point for musical applications, as we will shortly see.

For those familiar with calculus, the mathematical formulation of electromagnetic induction says that if the size of the field in teslas inside a coil with N turns is B, then the voltage V generated is

$$V = -NA\frac{\mathrm{d}B}{\mathrm{d}t}, \tag{7.40}$$

where A is the cross-sectional area of the coil. The quantity $\Phi = BA$ is called the *magnetic flux* in the coil, so you may also see the equation written as

$$V = -N\frac{\mathrm{d}\Phi}{\mathrm{d}t}. \tag{7.41}$$

Electromagnetic induction is of central importance in sound recording and reproduction. For instance, a dynamic microphone works like the reverse of a loudspeaker, with a coil attached to a diaphragm and a permanent magnet mounted nearby. When sound strikes the diaphragm it causes it to vibrate, moving the coil, which changes the magnetic field inside the coil and hence generates an electrical signal. We study microphones starting in the following section, and dynamic microphones in Section 7.3.6.

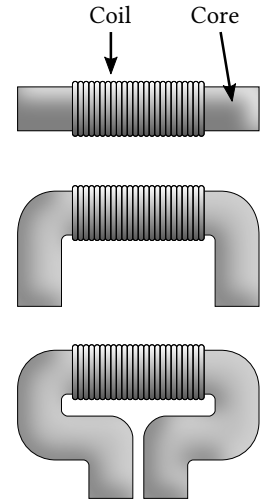


Figure 7.20: The core of a solenoid can be extended beyond the ends of the coil or bent around to carry the magnetic field in particular directions. This trick is used for instance in the magnetic "head" of a tape recorder.

7.3 Microphones

Armed with our knowledge of electrical signals and circuits, let us now turn our attention to how music can be captured electrically, recorded, and then reconstituted as sound again. The first step of the process, turning sound into electrical signals, is accomplished with a microphone, a device that detects sound and produces a varying electrical voltage that mirrors the sound waveform.[3] The basic scientific principle behind all microphones is the same: a thin membrane or *diaphragm* moves in response to the incoming sound in a manner similar to the movement of the eardrum in the ear (see Section 5.2.2), then this movement is translated into a corresponding electrical voltage via one of several mechanisms that we will describe shortly.

[3]In Chapter 15 we discuss electronic instruments, such as synthesizers, which directly generate an electrical signal without first producing any sound. For such instruments there is no need to use a microphone—we already have our electrical signal. For traditional musical instruments, however, including string and wind instruments, percussion, and voice, sound must be first captured with a microphone. A few instruments, most notably the electric guitar, are intermediate cases. The electric guitar directly produces an electric signal and so in principle no microphone is needed, but in practice the instrument is usually played through a guitar amplifier that produces sound and then the sound is captured with a microphone.

7.3.1 Pressure microphones

The most fundamental type of microphone is the *pressure microphone*, which directly captures sound pressure. It works by combining a diaphragm with a sealed *capsule*, a small container of air. The arrangement is shown in Fig. 7.21.

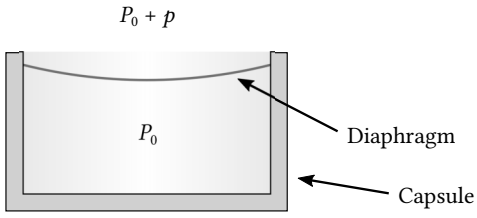


Figure 7.21: A pressure microphone. In a pressure microphone the capsule is sealed so that sound only reaches the outside of the diaphragm (the top in this figure) making the outside pressure $P_0 + p$ different from the inside pressure P_0 and causing the diaphragm to bow in or out.

The diaphragm is a thin, flexible membrane, normally made of mylar, a durable plastic. The diaphragm, which is usually circular, forms one wall of the capsule, and the remaining walls are made of rigid plastic or metal. The diaphragm is stretched tightly across the open end of the capsule, sealing it off from the outside air.

When sound arrives at the microphone it causes a change in air pressure. Recall that air pressure P is equal to the prevailing atmospheric pressure P_0 plus the sound pressure p. Inside the capsule, however, the pressure is just P_0, since the capsule is sealed off. The result is that the pressure differs on the outside and inside of the diaphragm. If p is positive, meaning the pressure is greater outside the capsule, it pushes on the diaphragm causing it to bow inward as shown in Fig. 7.21. If p is negative the pressure is greater on the inside of the capsule and causes the diaphragm to bow outward. In other words, the movement of the diaphragm reflects the net force acting on it.

If the pressure on the front of the diaphragm is P_f and its area is S, then the force on the front is $F_f = P_f S$. Similarly the force on the back is $F_b = P_b S$, and the net force F is the difference of the two:

$$F = F_f - F_b = P_f S - P_b S = (P_f - P_b)S. \tag{7.42}$$

In this case we have $P_f = P_0 + p$ and $P_b = P_0$, so

$$P_f - P_b = P_0 + p - P_0 = p, \tag{7.43}$$

and the force is simply

$$F = pS, \tag{7.44}$$

proportional to the sound pressure. The force is also proportional to the area S of the diaphragm, meaning that a larger diaphragm will feel a larger force. High-quality microphones typically have a large diaphragm for this reason, to make them more sensitive to quiet sounds. There are limits to how large the diaphragm can be however, for several reasons. Larger diaphragms are more expensive and they are also fragile and therefore may not be suitable for rougher conditions such as live stage

performance. The main limit on the size of the diaphragm, however, is that the capsule needs to be smaller than the wavelength of the sound if the microphone is to operate properly. This point is discussed further below.

The diaphragm is light enough that it responds very rapidly to changes in pressure. This means in practice that the position of the diaphragm, whether it is bowed inward or outward, just depends on the net force, i.e., on the sound pressure. If we can measure this position then we have in effect measured the sound pressure. There are two commonly used methods for measuring the position, using electrical capacitance and electromagnetic induction. We discuss these in Sections 7.3.5 and 7.3.6.

An important point about the pressure microphone is that, since it measures pressure, it is equally sensitive to sound coming from all directions. Even though sound travels in a particular direction, pressure does not have a direction. It just asks on whatever surface is presented to it, in this case the surface of the diaphragm (see the discussion on page 2). Thus a pressure microphone is *omnidirectional*. In practice a pressure microphone may be slightly more sensitive to sound coming from in front than from behind, because sound from behind is blocked in part by the body of the capsule itself. This, however, is a small effect. As discussed in Section 3.3, sound naturally flows around objects smaller than the sound wavelength and does not feel the obstruction. So as long as the microphone capsule is smaller than the wavelengths of the sounds it is capturing, sound will therefore flow around it. The shortest sound wavelengths, those of the highest frequency sounds, are about 2 cm, so the capsules of pressure microphones are deliberately designed to be smaller than this so that they pick up sound from all directions without obstruction.

For more discussion of this point see Example 3.6 on page 70.

The ability of a pressure microphone to pick up sound from any direction can be useful. Consider for instance the microphone in a mobile phone. It is important for your phone to pick up your voice clearly no matter how you hold it, or even if it is lying on the table or on the other side of the room. Phones use pressure microphones to achieve this. Pressure microphones are also used for headset and lapel microphones and for live recording, particularly when we want to capture many instruments at once (such as with an orchestra) or when we want to capture the acoustics of the performance space as well as the performance itself.

Another use of pressure microphones is in measuring the loudness of sounds using a sound level meter, as described in Section 3.7. Sound meters measure sound pressure level as defined in Eq. (3.22) on page 63, which requires us to measure the mean-square pressure over a period of time. To do this a sound meter uses a pressure microphone, directly measuring the pressure, then squares the results and averages them, usually over a period of one second, to calculate the mean-square pressure. This also means that sound meters are omnidirectional, measuring sound from all directions and not just the direction they are pointing in, which in most cases is a desirable feature.

7.3.2 Pressure-gradient microphones

A pressure-gradient microphone is also sometimes called a *velocity microphone*.

The alternative to a pressure microphone is a *pressure-gradient microphone*. The crucial difference between the two is that the capsule of a pressure-gradient microphone is not sealed, meaning that sound reaches both sides of the diaphragm. The arrangement is shown in Fig. 7.22. The capsule is now a cylindrical tube, closed at one end by the diaphragm but open at the other end. Like the pressure microphone, the diaphragm of a pressure-gradient microphone will bow one way or the other, depending on which side has the higher pressure.

A defining feature of the pressure-gradient microphone is that, unlike the omnidirectional pressure microphone, its response depends on the direction the sound comes from. Consider a sound wave arriving at the microphone not straight on but at an angle θ, as shown in Fig. 7.22. While sound can reach both top and bottom of the diaphragm, it has further to travel to get to the bottom. Sound going to the bottom has to travel the extra distance marked x in the figure to reach the far end of the capsule and then back up the length of capsule to the diaphragm, which is the distance marked d. From the geometry we can see that $x = d\cos\theta$, so the total extra distance traveled by the sound going to the back of the diaphragm is

$$x + d = d\cos\theta + d = d(\cos\theta + 1). \tag{7.45}$$

Suppose the incoming sound is a pure sine wave at frequency f. Then, from Eq. (4.1), the sound pressure at the front of the diaphragm can be written as

$$p_f(t) = A\sin(2\pi f t), \tag{7.46}$$

where A is the amplitude of the waveform, and the sound pressure at the back of the diaphragm is the same, but delayed by the amount of time Δt it takes to traverse the

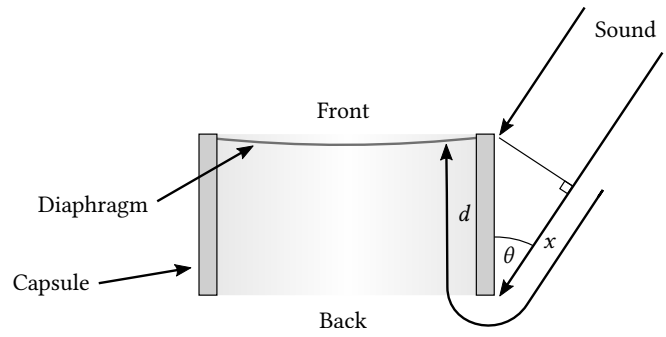


Figure 7.22: A pressure-gradient microphone. In a pressure gradient microphone the capsule is open at the back so that sound reaches both sides of the diaphragm, but the sound has to travel a longer distance, as marked, to reach bottom of the diaphragm.

extra distance, which is

$$\Delta t = \frac{d(\cos\theta + 1)}{c}, \tag{7.47}$$

where c is the speed of sound as usual. Thus the sound pressure at the back of the diaphragm is

$$p_b(t) = A\sin[2\pi f(t - \Delta t)]. \tag{7.48}$$

The total pressure on the front of the diaphragm is thus $P_f = P_0 + p_f$ and the pressure on the back is $P_b = P_0 + p_b$ and the difference of the two is

$$\begin{aligned} P_f - P_b &= (P_0 + p_f) - (P_0 + p_b) = p_f - p_b \\ &= A\sin(2\pi f t) - A\sin[2\pi f(t - \Delta t)] = 2A\sin(\pi f \Delta t)\cos[\pi f(2t - \Delta t)], \end{aligned} \tag{7.49}$$

where we have used a trigonometric identity to get the last formula. Comparing with Eq. (7.43), this result tells us that the pressure-gradient microphone will behave as if it were a *pressure* microphone with the effective sound pressure

$$p(t) = 2A\sin(\pi f \Delta t)\cos[\pi f(2t - \Delta t)], \tag{7.50}$$

where Δt is given by Eq. (7.47).

Now note that the quantity

$$f\Delta t = f\,\frac{d(\cos\theta + 1)}{c} = \frac{d}{\lambda}(\cos\theta + 1), \tag{7.51}$$

where $\lambda = c/f$ is the wavelength of the sound from Eq. (2.4). A typical size for a microphone capsule is about $d = 1$ or 2 cm, much smaller than the wavelength of all but the highest-frequency audible sounds, so d/λ is a normally small number, much less than one. Note also that $\sin\phi \simeq \phi$ when ϕ is a small number, so $\sin(\pi f \Delta t) \simeq (fd/c)(\cos\theta + 1)$ and to a good approximation we can write the effective sound pressure of Eq. (7.50) as

$$p(t) = \underbrace{\frac{2\pi A f d}{c}(\cos\theta + 1)}_{\text{Amplitude}}\ \underbrace{\cos(2\pi f t)}_{\text{Waveform}}, \tag{7.52}$$

where we have approximated the sine term and dropped the Δt term from the second cosine, since it's small and doesn't make much difference.

Recalling that the cosine wave $\cos(2\pi f t)$ looks just like a sine wave with frequency f except that it is shifted in time, this equation tells us that the pressure-gradient microphone will feel an effective sound pressure $p(t)$ in the shape of a sine wave with frequency f, but with amplitude $(2\pi A f d/c)(\cos\theta + 1)$. Because this amplitude is proportional to $\cos\theta + 1$, the effective pressure varies with the direction the

sound is coming from. By contrast, the pressure felt by a pressure microphone does not depend on direction—a pressure microphone is omnidirectional, as discussed in Section 7.3.1. Figure 7.23 shows a plot of $\cos\theta + 1$ and, as we can see, the value peaks at 0°, directly in front of the microphone, and falls all the way to zero at 180°. This is a *directional microphone*. It picks up sound from the direction you point it in and no sound at all from behind. A microphone with this particular directional pattern is called a *cardioid microphone*.

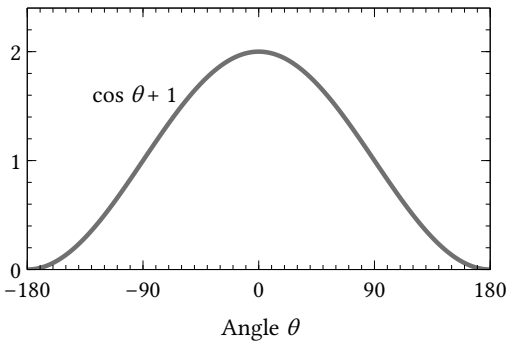


Figure 7.23: The amplitude of the effective sound pressure felt by a cardioid microphone. The pressure felt by a cardioid microphone is greatest when the sound comes from immediately in front of the microphone (0°) and zero when it comes from behind (180° or −180°). The equivalent polar plot is shown in Fig. 7.25.

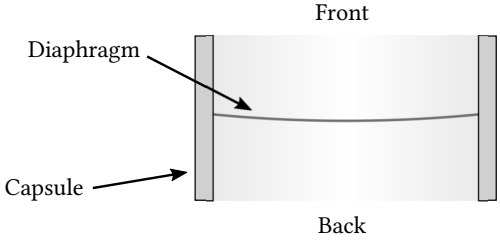


Figure 7.24: A bidirectional pressure-gradient microphone. If the diaphragm of a pressure-gradient microphone is mounted in the center of the capsule it produces a bidirectional response that is strongest at the back and front of the microphone and weakest at the sides—see Fig. 7.25.

This directionality is very convenient for many musical applications, for instance in live performance situations. You can point the microphone towards a performer or instrument you want to capture and it will pick up mostly their sound and relatively little sound from other directions, such as from other performers, on-stage speakers, or audience members. In practice, real directional microphones may not exactly match the theory—they might pick up a little sound from behind for instance—but they can definitely cut down a lot on unwanted sounds.

A variant on the cardioid microphone is one in which the diaphragm is placed not at the top of the capsule but in the middle, as shown in Fig. 7.24. Now the sound at the front of the capsule has to travel a distance $\frac{1}{2}d$ down the length of the capsule to reach the diaphragm, while the sound at the back has to travel $x + \frac{1}{2}d$, with x as in Fig. 7.22. Thus the sound going to the back of the microphone needs to travel further than the sound going to the front by a distance $(x + \frac{1}{2}d) - \frac{1}{2}d = x = d\cos\theta$. Modifying Eqs. (7.50) and (7.52) appropriately, we find that the effective pressure felt by the microphone is now

$$p(t) = \underbrace{\frac{2\pi A f d}{c}\cos\theta}_{\text{Amplitude}}\ \underbrace{\cos(2\pi f t)}_{\text{Waveform}}. \qquad (7.53)$$

So the amplitude is proportional to $\cos\theta$, which takes its largest values at $\theta = 0°$ and 180° and is zero at ±90°. In other words, the microphone now picks up sound strongly from directly in front and behind, but not at all from the sides. This is a *bidirectional* microphone.

A third common choice is one in which the diaphragm is placed $\frac{2}{3}$ of the way up the capsule, producing

a *hypercardioid* microphone, which is a compromise between cardioid and bidirectional. It has a narrower pickup pattern at the front than a true cardioid, which is good if one wants a highly directional microphone, but it also picks up a small amount of sound from behind.

7.3.3 Directional response

The directionality of microphones is often depicted using a *polar plot*, a circular type of graph as shown in Fig. 7.25. The direction the sound is coming from is represented by position around the circle, starting at the top which represents sound coming from straight in front of the microphone. The line on the plot represents how sensitive the microphone is in the corresponding direction in terms of the amplitude of the signal, with values further from the center denoting higher amplitude. In the first plot of Fig. 7.25, for example, the line is equally far from the center the whole way around, indicating an omnidirectional microphone—the classic pressure microphone. The second plot shows a cardioid microphone, whose sensitivity is greatest at the front and goes to zero at the back, at 180°. The shape of this second plot is said by some people to look like an upside-down heart and this is the origin of the name "cardioid," which means "like a heart." The third plot shows the response pattern of a hypercardioid microphone and the fourth shows a bidirectional microphone, sometimes also called a "figure 8" microphone because the polar plot looks like an 8.

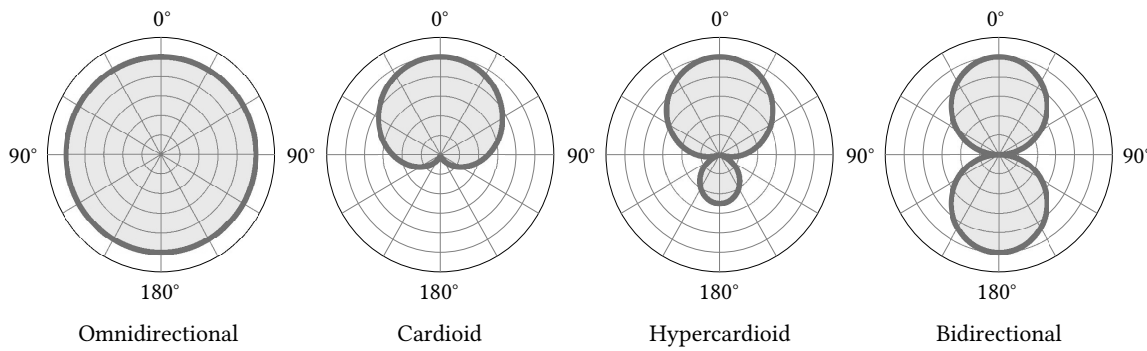


Figure 7.25: Polar plots of the directional sensitivity of microphones. In a polar plot the direction of the incoming sound is measured around the circle, with the top of the plot representing the front of the microphone. Distance of the line from the center of the plot represents the amplitude of the response of the microphone to sound from the corresponding direction. Thus, in the omnidirectional case the line is equally far from the center all the way around, making a circle on the plot, so the microphone is equally sensitive in all directions. The cardioid microphone is sensitive mainly to sound from the front, but not the sides and back. In the bidirectional case the microphone is sensitive to sound from the front and the back, but not the sides, while the hypercardioid falls in between the cardioid and bidirectional, being sensitive mostly to sound from the front and somewhat less from the back.

EXAMPLE 7.3: DIRECTIONAL MICROPHONES

How much smaller is the intensity recorded by a cardioid microphone if a sound hits it 90° off-center versus straight-on?

To answer this question, we look at Eq. (7.52). The dependence of the effective sound pressure on the angle θ that the sound is coming from is given by the factor $\cos\theta + 1$. For sound hitting the microphone from the front, at $\theta = 0$, this factor is $\cos 0 + 1 = 2$, but for sound at 90° it is $\cos 90° + 1 = 1$. So the sound pressure at the side is a half what it is at the front. Recall, however, that intensity is proportional to sound pressure squared (see Eq. (3.6) on page 57), so the intensity at 90° is a quarter what it is at 0°, or 6 dB lower.

7.3.4 FREQUENCY RESPONSE

A point to notice about Eq. (7.52) for the pressure-gradient microphone is that the amplitude of the effective sound pressure $p(t)$ is proportional to the frequency f. This means that the pressure-gradient microphone will register a stronger signal at high frequencies than it does at low ones, in contrast with a pressure microphone, for which $p(t)$ has no variation with frequency. Recall that the intensity of a sound is proportional to pressure squared (see Eq. (3.7) on page 57), so the effective intensity felt by a pressure-gradient microphone will be proportional to f^2. All other things being the same, this means the intensity will go up by a factor of four, or 6 dB, when the frequency goes up by a factor of two, i.e., an octave. This makes the sound captured by a pressure-gradient microphone brighter, with more treble, than a comparable pressure microphone.

At the highest frequencies, the approximation we made in Eq. (7.52), that $\sin\phi \simeq \phi$, is not accurate and we need to calculate the amplitude of the sound pressure from the full expression in Eq. (7.50). Putting this expression into the formula for intensity, Eq. (3.9), we get

$$I = \frac{\langle p^2\rangle}{\rho c} = \frac{4A^2}{\rho c}\sin^2(\pi f\Delta t)\bigl\langle\cos^2[\pi f(2t-\Delta t)]\bigr\rangle = \frac{2A^2}{\rho c}\sin^2\Bigl(\frac{\pi f d}{c}(\cos\theta+1)\Bigr), \quad (7.54)$$

where we have made use of the fact that the average of $\cos^2\phi$ is $\frac{1}{2}$.

Converting to decibels, we show in Fig. 7.26 a plot of the resulting sound level as a function of frequency for sound coming from directly in front of the microphone at $\theta = 0°$ and a capsule of depth $d = 2$ cm. The 6 dB-per-octave increase in sensitivity is clearly visible in the lower frequencies of the plot, but at higher frequencies the response levels off and has some pronounced dips. These dips correspond to points where the sine function in (7.54) is zero. The first zero of the sine function, for instance, occurs when the argument of the function is equal to π, or in other words when

$$f = \frac{c}{d(\cos\theta + 1)}. \quad (7.55)$$

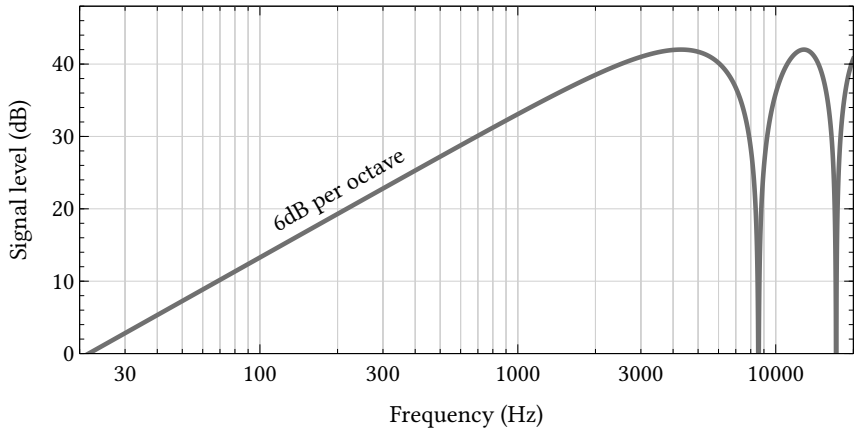


Figure 7.26: Theoretical frequency response of a pressure-gradient microphone. A pressure-gradient microphone is more sensitive to higher frequency sounds. The effective intensity felt by the microphone is given by Eq. (7.54), which tells us that the sound level goes up by 6 dB per octave over most of the frequency range, but with marked dips at high frequencies. Real microphones use a variety of techniques to smooth out this behavior, as discussed in the text.

For the angle $\theta = 0°$ used in the figure (for which this frequency is lowest), and putting $d = 0.02\,\text{m}$ and $c = 343\,\text{m/s}$, the dip falls at $f = 8575\,\text{Hz}$, well within the human hearing range.

The frequency response shown in Fig. 7.26 is less than ideal, for two reasons. First, while an increased sensitivity to high frequencies is desirable in some cases, the increase of 6 dB per octave is too large. There is a variation of over 40 dB between the lowest and highest frequencies shown in the figure, a factor of 10 000 in intensity, which will be very noticeable. Second, the behavior at high frequencies, with dips in the response, is definitely undesirable: it means that the microphone will be uneven in the upper range and will fail to pick up some frequencies altogether.

A number of techniques are used to mitigate these problems. The overall increase in response can be reduced or eliminated by passing the signal from the microphone through a low-pass filter of the kind described in Section 7.2.4. Alternatively, one can reduce the high-frequency response mechanically either by increasing the mass of the diaphragm, which makes it slower to respond to high frequencies, or by increasing the "damping." For the interested reader, the details are explained in Section 7.3.7.

To compensate for the high-frequency dips in Fig. 7.26 one can reduce the depth d of the microphone capsule, which doesn't eliminate the dips but pushes their frequency up, via Eq. (7.55), ideally to the point where their effect can no longer be

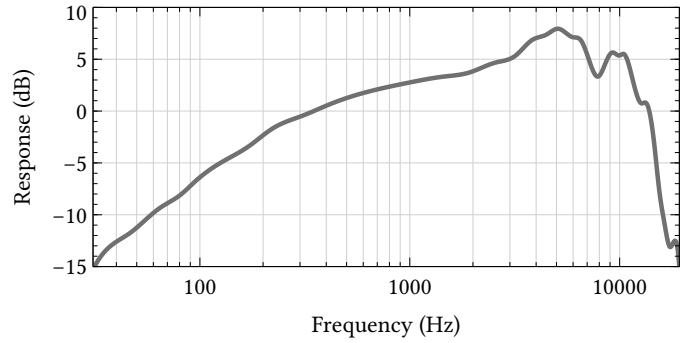


Figure 7.27: Frequency response of a typical cardioid microphone. This pressure-gradient microphone shows an increase in sensitivity with frequency at the lowest frequencies, but the response levels off as frequency gets higher because the microphone has a heavier diaphragm that limits the high-frequency response. At the highest frequencies the sensitivity drops off sharply, which may be because of dips in the response of the kind shown in Fig. 7.26, or because of the mass of the diaphragm or limitations of the electronics.

heard. Decreasing d, however, also decreases the overall amplitude of the signal via Eq. (7.52), so the choice of d is necessarily a compromise. The smallest capsules are about 1 cm deep, which raises the frequency of the first dip to 17 150 Hz, right at the upper frequency limit of audibility, but some microphones opt for a deeper capsule that produces a stronger signal at the expense of a less satisfactory response at the highest frequencies.

Figure 7.27 shows the real-world measured frequency response of a pressure-gradient microphone. As the figure shows, the sensitivity of the microphone does increase with frequency, particularly at lower frequencies, but then levels off as frequency gets higher. In this case this is achieved by using a heavier diaphragm: the mylar film of the circular diaphragm of this microphone is thicker in the middle to increase its mass but thinner around the edge to allow it to move freely with the incoming sound waves.

A further issue with pressure-gradient microphones is that the directional response at the highest frequencies, given by Eq. (7.54), is no longer simply proportional to $\cos\theta + 1$, and hence does not follow the standard cardioid shape. In practice, this means that pressure-gradient microphones tend to become less directional at high frequency. Figure 7.28 shows the directional response of the microphone from Fig. 7.27 at three different frequencies, and we can see that the microphone has a true cardioid response only at the lowest of the three and becomes more omnidirectional at higher frequencies.

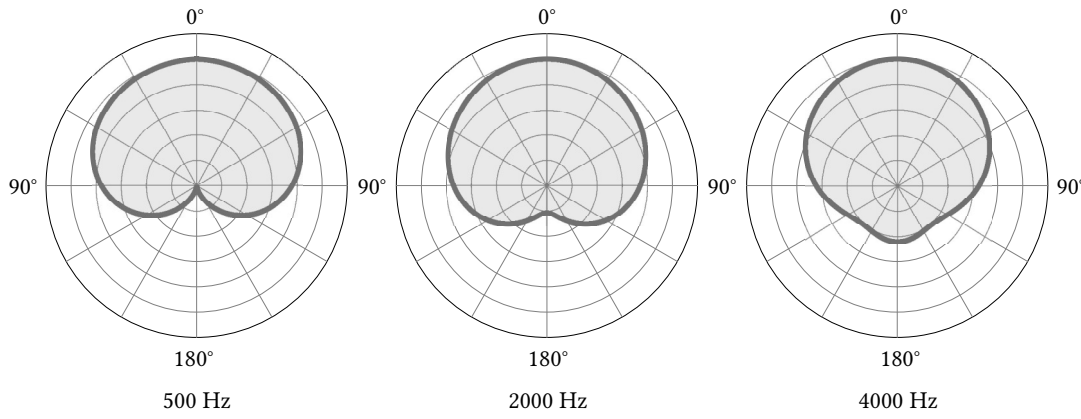


Figure 7.28: Directionality of a typical cardioid microphone. These polar plots show the measured directional response of the same microphone as in Fig. 7.27 at three different frequencies as indicated. In principle, this is a cardioid microphone, and it does show the classic cardioid at 500 Hz, but it becomes progressively less directional as frequency increases.

7.3.5 Condenser microphones

So far we have discussed how the diaphragm of a microphone moves in response to incoming sound. To make a functioning microphone, however, we also need some way of detecting this movement and turning it into an electrical signal. There are two primary means by which this is accomplished. The first makes use of the phenomenon of electrical capacitance. The second makes use of electromagnetic induction. We will look at each of these in turn, starting in this section with capacitance.

A condenser microphone

As discussed in Section 7.2.3, an electrical capacitor is a device made of two metal plates separated by a small gap. When a voltage is applied between the plates a small amount of electrical charge—electrons—flows onto them and stays there. The amount of charge Q is given by the capacitor equation $Q = CV$ where V is the voltage across the capacitor and C is the capacitance (see Eq. (7.12)). The value of the capacitance depends on the construction of the capacitor and particularly on the distance between the plates. If the distance is z then the capacitance is inversely proportional to z thus:

$$C = \frac{k}{z}, \tag{7.56}$$

where k is a constant. So when the distance is smaller the capacitance is bigger. We can use this effect to detect the movement of the diaphragm in a microphone. A microphone of this kind is called a *condenser microphone*.

Figure 7.29 shows how such a microphone works. The diaphragm acts as one plate of a capacitor and we add another fixed plate next to it called the *backplate*. As

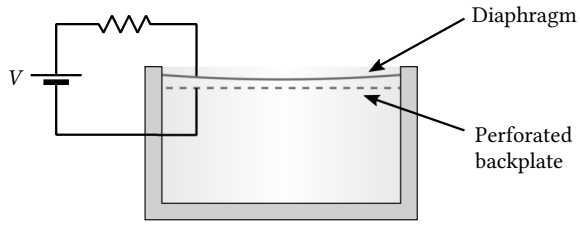


Figure 7.29: A condenser microphone. A condenser microphone has a diaphragm coated in metal to make it conduct and a metal backplate, perforated to allow sound and air to pass through it. Together the diaphragm and backplate form a capacitor whose capacitance depends on the distance between the two and hence varies as the diaphragm moves.

the diaphragm moves, the distance between the two plates varies and hence so does the capacitance. We use a power supply to place a fixed charge on the capacitor and as the capacitance varies this charge produces a varying voltage. All we have to do is measure this voltage and we have our electrical signal.

This is the basic principle, but there are some important details. First, the diaphragm is normally made of mylar plastic, which doesn't conduct electricity, so it cannot be directly used as a capacitor plate. To get around this problem, the diaphragm is coated with a thin, flexible layer of metal, most often gold, to make it conduct. If you look closely at a condenser microphone you can sometimes see the gold color of the diaphragm behind the protective grille on the outside of the microphone.

The backplate is made of rigid metal and is normally placed immediately behind the diaphragm, out of sight and just a small distance from the diaphragm, typically only a few micrometers. If the backplate were completely solid it would create a back-pressure that would hinder the movement of the diaphragm, and for pressure-gradient microphones it would also block sound from reaching the back of the diaphragm, which is important for the way these microphones operate. To get around these issues, the backplate is normally perforated with a set of holes to let air and sound through.

This is the circuit symbol for a microphone

The basic circuit diagram for a condenser microphone is shown in Fig. 7.30a. The microphone is wired in series with a resistor and a constant voltage V from a battery or power supply is applied across both microphone and resistor. As described in Section 7.2.3, this causes current to flow through the resistor and charge up the capacitor (i.e., the microphone) until the voltage across it reaches V, at which point the current stops and the charge holds steady. If there is no sound arriving, so that the diaphragm is not moving, and if the capacitance of the microphone is C_0 when the diaphragm is at rest, then the capacitor law of Eq. (7.12) tells us that the charge

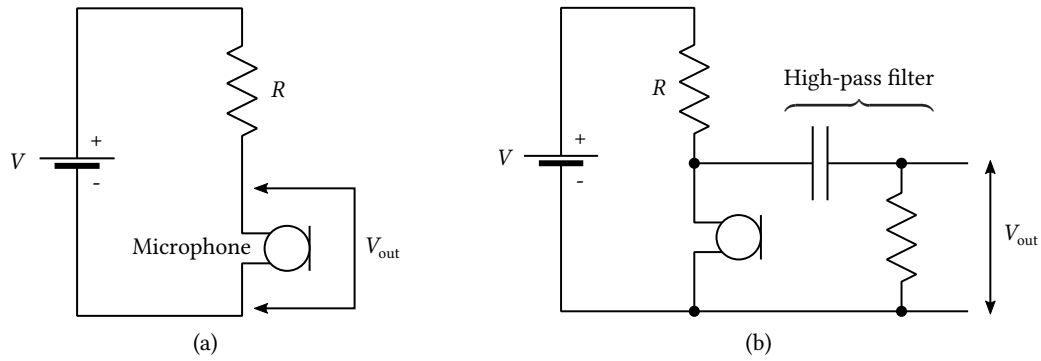


Figure 7.30: Circuit diagram for a condenser microphone. (a) A condenser microphone (conventionally represented by the symbol with a circle and a line) is wired in series with a large resistor R and a power supply with voltage V, producing a constant charge on the microphone. The varying capacitance of the microphone then causes the voltage V_{out} across it to vary according to Eq. (7.60). (b) By adding a high-pass filter to the output of the circuit, consisting of a capacitor and a second resistor (see Section 7.2.4), we can remove the constant voltage V, leaving us with the output voltage given in Eq. (7.61).

on the microphone will be

$$Q = C_0 V. \tag{7.57}$$

The resistance of the resistor is chosen to be very large—typically billions of ohms—so even the largest current $I = V/R$ that can flow through it is very small. This means that it takes a while for the capacitor to charge up, but once it does the charge remains essentially constant thereafter, with the value Q from Eq. (7.57). Even when sound falls on the diaphragm and pushes it one way or the other, at this slow charging rate there simply isn't enough time for the charge to change much before the diaphragm moves back again.

See Exercise 7.9 on page 279 for more details on how this works.

What does change when sound arrives however is the capacitance C, which varies with the separation between the plates of the capacitor, i.e., between the diaphragm and backplate. With the charge remaining constant, this means that the output voltage V_{out} across the microphone will be

$$V_{\text{out}} = \frac{Q}{C} = \frac{C_0 V}{k} z, \tag{7.58}$$

where we have used Eqs. (7.56) and (7.57). Let us write the distance between the diaphragm and the backplate as $z = z_0 + \Delta z$, where z_0 is the distance when the diaphragm is at rest and Δz is how much it moves due to the sound pressure. Then

$$V_{\text{out}} = \frac{C_0 V}{k}(z_0 + \Delta z). \tag{7.59}$$

But by definition the capacitance at rest is $C_0 = k/z_0$ and so

$$V_{\text{out}} = \frac{V}{z_0}(z_0 + \Delta z) = V + \frac{V}{z_0}\Delta z. \tag{7.60}$$

The constant voltage V can be filtered out using a high-pass filter with a low cutoff frequency, as described in Section 7.2.4. This gives us the circuit shown in Fig. 7.30b, with output voltage

$$V_{\text{out}} = \frac{V}{z_0}\,\Delta z. \tag{7.61}$$

So the output voltage simply mirrors the displacement Δz of the diaphragm.

Equation (7.61) is inversely proportional to z_0, the separation at rest between the diaphragm and the backplate, so by making this separation small we can make the output signal large. We should, however, avoid making the separation so small that diaphragm and backplate might touch when the diaphragm moves, which would cause a short-circuit and destroy the charge on the capacitor. 20 to 40 micrometers is a typical separation distance for commercial condenser microphones.

Equation (7.61) is also proportional to the power-supply voltage V, which means we can increase the signal by increasing V, though we want to keep the voltage small enough to not risk shocking the user of the microphone if something goes wrong. By longstanding convention, condenser microphones use a voltage of 48 volts. This is too large to be conveniently supplied by a battery, so condenser microphones normally use an external power supply, known in the music industry as a "phantom" power supply.

Even when we take these steps to make the signal from the microphone as large as possible, it is still very small—typically just a fraction of a volt—so it is amplified with a special preamplifier designed for use with microphones. Commonly the preamplifier and the phantom power supply are combined into one device that both generates the voltage for the microphone and amplifies the resulting signal.

Condenser microphones can be made with large diaphragms that are highly sensitive to incoming sound and give very high quality recordings. They are widely used in recording studios. However, they can also be fragile and easily damaged, and the fact that they require a separate power supply makes them less convenient than other types of microphones. In live music situations, where ruggedness is a bonus, or in situations such as amplification of speech, where convenience is at a premium and sound quality is not, another type of microphone, the dynamic microphone, is more widely used.

Example 7.4: Condenser microphone design

A certain condenser microphone runs on the standard 48 volt power. An incoming sound causes the diaphragm to move back and forth by 0.05 μm each way. If we want the micro-

phone to produce a corresponding 0.1 volts in the output voltage what should we make spacing between the diaphragm and the backplate?

We can answer this question using Eq. (7.61). Rearranging the equation for the spacing z_0, we have

$$z_0 = \frac{V}{V_{\text{out}}}\,\Delta z = \frac{48}{0.1} \times 0.05\,\mu\text{m} = 24\,\mu\text{m}, \tag{7.62}$$

which is typical for today's condenser microphones.

7.3.6 Dynamic microphones

The primary alternative to a condenser microphone is a *dynamic microphone*, which uses electromagnetic induction to measure the movement of the diaphragm. Dynamic microphones can give results of good quality, but cannot match the sound from the best condenser microphones. They are popular, however, for their ruggedness, which makes them ideal for live performance situations.

A dynamic microphone

As discussed in Section 7.2.7, electromagnetic induction occurs when the magnetic field passing through a coil of wire changes, getting either stronger or weaker, which produces an electric voltage between the ends of the coil. We can use this effect to build a microphone. Perhaps the most obvious way to proceed would be to attach a magnet to a diaphragm and place a coil nearby. The movement of the diaphragm would then make the magnet move and hence change the magnetic field in the coil, producing a current. In practice, however, this approach doesn't work well because magnets are too heavy. It is important that the diaphragm be light, so it can respond rapidly to changes in pressure, so having a heavy magnet attached is not ideal.

Instead, therefore, dynamic microphones work the other way around, by attaching the coil (which is relatively light) to the diaphragm and placing a stationary magnet nearby—often actually inside the coil or partially inside it. Figure 7.31 shows a

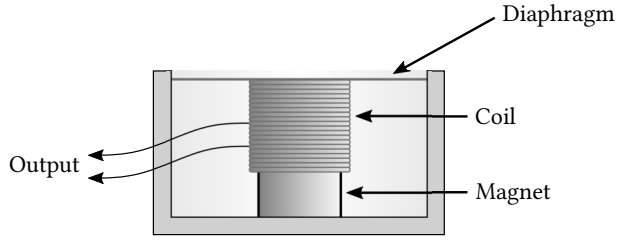


Figure 7.31: A dynamic microphone. In a dynamic microphone a coil is attached to the back of the diaphragm and moves as the diaphragm vibrates. A fixed permanent magnet is mounted so that it extends inside the coil. When the coil moves, the magnetic field running through it changes, producing a current in the coil which can then be amplified or recorded.

typical arrangement, with the diaphragm forming the top of the microphone capsule and the coil attached to its underside. The coil moves with the diaphragm, and hence the magnetic field inside it, produced by the stationary magnet, changes, inducing a current. The magnetic field itself doesn't move, but the amount of it inside the coil changes as the coil goes back and forth.

In addition to their ruggedness, dynamic microphones have an advantage over condenser microphones in not requiring an external power supply—the process of electromagnetic induction creates its own electricity. Dynamic microphones come in both pressure and pressure-gradient varieties. The most common type for musical applications are cardioid pressure-gradient microphones.

Advanced material

7.3.7 Mechanics of microphones

The signal produced by a microphone depends on a combination of the movement of the diaphragm in response to the sound pressure acting on it and the method used to turn the movement into electricity. The movement of the diaphragm is driven by the force due to sound pressure, Eq. (7.44), which we here denote by $F_p = Sp(t)$, where S is the area of the diaphragm and $p(t)$ is either the actual sound pressure (for a pressure microphone) or the effective pressure, Eq. (7.52), from the difference between the front and back (for a pressure-gradient microphone).

This is not the only force acting on the diaphragm however. There are also two other forces. The first is a restoring force F_r due to the elastic nature of the diaphragm. When the diaphragm is pushed to one side or the other its springiness naturally pulls it back. To a good approximation, the restoring force obeys Hooke's spring law. If the displacement of the diaphragm from its rest position is ξ, then the restoring force is $F_r = -k\xi$, where k is the spring constant and the minus sign indicates that the restoring force is in the opposite direction to the displacement.

The other force on the diaphragm is a damping force F_d that acts to slow down the diaphragm when it is moving. This force comes primarily from air resistance and is proportional to the velocity of the diaphragm $\mathrm{d}\xi/\mathrm{d}t$, so that $F_d = -\gamma\,\mathrm{d}\xi/\mathrm{d}t$, where the minus sign indicates that the force is in the opposite direction to the velocity and γ is a constant which controls the strength of the damping and which depends on the size of the diaphragm and on the shape and design of the microphone capsule.

Putting together the forces on the diaphragm, the total force is $F = F_p + F_r + F_d$. Applying Newton's second law in the form $F = ma$ and writing the acceleration as $a = \mathrm{d}^2\xi/\mathrm{d}t^2$, we then find that the motion of the diaphragm obeys the differential equation

$$m\frac{\mathrm{d}^2\xi}{\mathrm{d}t^2} + \gamma\frac{\mathrm{d}\xi}{\mathrm{d}t} + k\xi = Sp(t). \tag{7.63}$$

Suppose now that sound strikes the microphone in the form of a sine wave with frequency f, which we write in the standard manner as the real part of a complex wave $p(t) = p_0 \mathrm{e}^{\mathrm{i}\omega t}$, where $\omega = 2\pi f$ and p_0 is either equal to the amplitude A of the wave for a pressure microphone or, for a pressure-gradient microphone, is given by Eq. (7.52):

$$p_0 = \frac{\omega A d}{c}(\cos\theta + 1). \tag{7.64}$$

The sine-wave sound pressure produces a corresponding sine-wave motion in the diaphragm $\xi(t) = \xi_0 \mathrm{e}^{\mathrm{i}\omega t}$ and, substituting into Eq. (7.63) and cancelling a factor of $\mathrm{e}^{\mathrm{i}\omega t}$, we find that $-\omega^2 m\xi_0 + \mathrm{i}\omega\gamma\xi_0 + k\xi_0 = Sp_0$, or

$$\xi_0 = \frac{Sp_0}{k - \omega^2 m + \mathrm{i}\omega\gamma}. \tag{7.65}$$

How this motion gets turned into an electrical signal depends on whether we have a condenser microphone or a dynamic microphone. Let us look at each in turn.

Condenser microphone: For a condenser microphone, the output voltage V_{out} is given by Eq. (7.61) with $\Delta z = \xi(t)$:

$$V_{\text{out}} = \frac{V}{z}\xi(t), \tag{7.66}$$

with V being the power supply voltage and z being the spacing between the diaphragm and the backplate. The amplitude of the output is then given by the complex magnitude of V_{out}, which is

$$|V_{\text{out}}| = \frac{V}{z}|\xi_0| = \frac{VSp_0/z}{\sqrt{(k-\omega^2 m)^2+\omega^2\gamma^2}}. \tag{7.67}$$

If our microphone is to have a uniform frequency response, we want this magnitude to be simply proportional to the amplitude A of the incoming sound wave and independent of ω. In a pressure microphone, for which $p_0 = A$, this requires that the denominator of (7.67) itself be independent of ω, which we can achieve by making m and γ small, so that $\sqrt{(k-\omega^2 m)^2+\omega^2\gamma^2} \simeq k$ and hence $|V_{\text{out}}| = VSA/zk$. This means we want a very thin, light diaphragm that responds rapidly to changes in sound pressure.

On the other hand, if we have a pressure-gradient microphone, then $p_0 \propto \omega$ as in Eq. (7.64). In this case, we can do one of two things. Either we can again use a thin, light diaphragm so that the denominator in (7.67) is constant, which produces a $|V_{\text{out}}|$ that is proportional to frequency ω, but then filter the resulting signal using a 6 dB-per-octave low-pass filter to remove this frequency dependence. Or we can design the microphone capsule so that the damping γ is large compared to the mass m and spring constant k of the diaphragm, which means $\sqrt{(k-\omega^2 m)^2+\omega^2\gamma^2} \simeq \omega\gamma$ and

$$|V_{\text{out}}| = \frac{(\omega SAd/c)(\cos\theta+1)}{\omega\gamma} = \frac{SAd}{\gamma c}(\cos\theta+1). \tag{7.68}$$

We cannot make γ very large, however, because a large γ will reduce the overall value of $|V_{\text{out}}|$ and hence reduce the output of the microphone. This typically means we have to make a compromise: limiting the value of γ may mean that for very low frequencies the denominator of (7.67) is not exactly proportional to ω and we will have some residual frequency dependence in V_{out}.

Dynamic microphone: For a dynamic microphone the output voltage is produced by induction in the microphone coil and Eq. (7.41) tells us that the voltage across a coil with N turns is

$$V_{\text{out}} = -N\frac{\mathrm{d}\Phi}{\mathrm{d}t}, \tag{7.69}$$

where Φ is the magnetic flux, i.e., the amount of magnetic field that passes through the coil. In this case the flux is proportional to the displacement ξ of the diaphragm, so

$$V_{\text{out}} = -Nb\frac{\mathrm{d}\xi}{\mathrm{d}t} = -Nb \times \mathrm{i}\omega\xi_0 \mathrm{e}^{\mathrm{i}\omega t}, \tag{7.70}$$

where b is a proportionality constant. With ξ_0 as in Eq. (7.65), the amplitude of the output voltage is then again given by the complex magnitude of V_{out}, which is

$$|V_{\text{out}}| = \frac{NbSp_0\omega}{\sqrt{(k-\omega^2 m)^2+\omega^2\gamma^2}}. \tag{7.71}$$

A crucial difference from the condenser microphone is the appearance of ω in the numerator of this expression, which means the dynamic microphone will produce a stronger signal at high frequencies. This is a result of the fact that electromagnetic induction responds to the rate of change of the magnetic field, and not just the size of the field, and it is separate from, and in addition to, the effect discussed in Section 7.3.4 where pressure-gradient microphones register a stronger signal at higher frequencies. We need to compensate for this effect if we want our microphone to have a uniform frequency response. How we do this depends again on whether we have a pressure microphone or a pressure-gradient microphone.

For a pressure microphone p_0 is simply equal to the amplitude A of the incoming sound wave, so the numerator of (7.71) is proportional to ω. In a manner similar to the pressure-gradient condenser mic, we can then achieve a uniform frequency response either by using a light diaphragm with small mass m and damping γ, which gives $|V_{\text{out}}| = NbSp_0\omega/k$, and then passing the output signal through a 6 dB-per-octave low-pass filter to remove the frequency dependence, or we can make the damping γ large, which gives $|V_{\text{out}}| = NbSp_0/\gamma$, which is independent of frequency from the outset.

Conversely, if we have a pressure-gradient dynamic microphone, with p_0 as in Eq. (7.64), then

$$|V_{\text{out}}| = \frac{(Nb\omega^2 SAd/c)(\cos\theta+1)}{\sqrt{(k-\omega^2 m)^2+\omega^2\gamma^2}}, \tag{7.72}$$

with a numerator that now goes as ω^2. We can also combat this frequency dependence in a number of ways. The traditional approach is to make a "mass-controlled" microphone with a heavier diaphragm, as described in Section 7.3.4, for which m is large compared to k and γ, which means $\sqrt{(k-\omega^2 m)^2+\omega^2\gamma^2} \simeq \omega^2 m$ and

$$|V_{\text{out}}| = \frac{NbSAd}{mc}(\cos\theta+1). \tag{7.73}$$

We should not make the mass too large, however, because it will reduce the overall value of $|V_{\text{out}}|$, and again this usually means we have to make a compromise: limiting the value of m may mean that for very low frequencies the denominator of (7.72) is not exactly proportional to ω^2 and we will have some frequency dependence in V_{out}. This kind of behavior can be seen for instance in Fig. 7.27 on page 244, where the response of the microphone does vary at low frequencies but then levels off.

Alternatively for this type of microphone we can use a lighter diaphragm and then pass the output through a low-pass filter, either cutting 6 dB per octave if the motion of the diaphragm is dominated by the damping γ or 12 dB per octave if both the mass and the damping are small.

Finally, note that for all variants of the dynamic microphone the output voltage is proportional to the number of turns N in the coil, so it pays to have as many turns as possible. There is a limit to how many we can have if we want to keep the mass of the coil reasonably low, but a typical microphone coil has a few thousand turns.

7.3.8 Other types of microphone

Condenser and dynamic microphones are the primary types of microphones used in modern music recording and amplification, but there are other types that see occasional use.

An *electret microphone* is a variant of a condenser microphone that uses a backplate with a permanent electric charge. Recall that a condenser microphone works by forming a capacitor between the diaphragm and the backplate, then applying a voltage across this capacitor using a power supply in order to place a charge on it. In an electret microphone one dispenses with the applied voltage and instead makes the backplate out of a special material—an electret—that carries a permanent electric charge embedded in it. This charge behaves in the same way as the applied charge in a conventional condenser microphone, producing a voltage across the capacitor that varies with the distance between diaphragm and backplate.

Compared to condenser microphones, electret microphones have the advantage of not requiring an external power supply, but their sound quality is often not as good as true condenser microphones, making them less popular for musical applications. They do find wide use however for recording speech, in dictation machines for example, where sound quality is not so important but convenience is.

A *ribbon microphone* is a variant of a dynamic microphone in which the coil of Fig. 7.31 is replaced by a thin ribbon of metal foil that acts both as diaphragm and as coil. The ribbon is mounted in the magnetic field of a permanent magnet and when sound hits the ribbon and causes it to move, it induces an electric current that can then be fed into an amplifier. Ribbon microphones are simple and durable, and hence attractive for live performance, but they produce only a small electric signal compared to a conventional dynamic microphone and normally have a bidirectional pickup pattern (see Fig. 7.25), which is less desirable than the cardioid pattern favored for most live applications.

A *carbon microphone* uses a completely different mechanism to detect sound. As shown in Fig. 7.32, the capsule is filled with granules of carbon, which have the

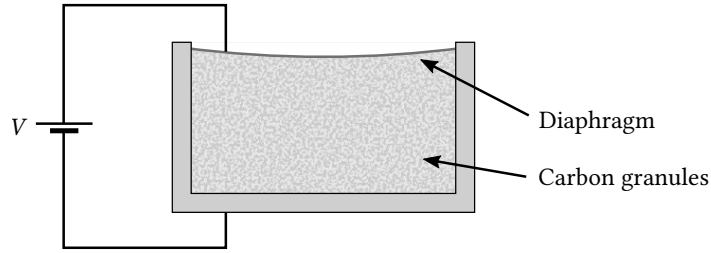


Figure 7.32: A carbon microphone. The capsule of a carbon microphone is filled with carbon granules whose resistance to the flow of electricity decreases when they are compressed. A voltage is applied as shown, producing an electric current through the carbon, and the motion of diaphragm squeezes the granules causing their resistance to vary and hence varying the current.

property that they conduct electricity better when compressed. The diaphragm of the microphone, which is typically made of thin metal foil, rests on the carbon granules, and a voltage is applied between the diaphragm and the bottom of the capsule, causing a current to flow through the carbon. When the diaphragm moves with incoming sound it compresses the carbon granules, causing them to conduct electricity better and so increasing the current. The result is a flowing current that varies with the sound waveform.

Carbon microphones are cheap, simple, robust, and do not require high voltages to operate, but they produce poor quality sound, not suitable for musical applications. The best known example of a carbon microphone for many people will be the mouthpiece in an old-fashioned landline telephone. Most telephones used carbon microphones until the 1980s, although modern landline phones and mobile phones typically use condenser or electret microphones.

A vintage telephone with a carbon microphone.

7.4 Recording

Once sound has been converted from a sound wave into an electrical signal it can be recorded using a variety of methods. For most of its history, sound recording was dominated by two main technologies, magnetic tape and vinyl records. Some of the most important recordings of all time have been made using these technologies, but both are now considered obsolete, having been superseded from the 1980s onward by digital recording techniques that convert sound waves into numbers which are then stored on hard disks, optical disks, or in the memory of computing devices. Digital music is a large topic that we deal with separately in the following chapter. In this section, we discuss tape and records. Although both technologies are obsolete it is worth taking a look at them in order to understand how they work, *why* they are

obsolete, and why modern digital recording technologies are superior.

7.4.1 Magnetic tape

A traditional reel-to-reel tape recorder

Throughout most of the twentieth century the principal medium for recording sound was magnetic tape, in the form of reel-to-reel tape recorders, studio multitrack recorders, cassettes, and 8-track tape. All of these operate on the same principle, making use of reels of flexible plastic tape coated with powdered ferric oxide, a naturally occurring mineral that contains high quantities of iron. When it comes into close contact with a magnetic field, the iron becomes magnetized—it becomes a weak magnet in its own right. The basic setup is shown in Fig. 7.33. A sound signal is fed into the coil of an electromagnet, called a *tape head*, with a specially shaped core designed to concentrate the magnetic field in a small region between its ends or *poles*. The tape passes over this region and becomes magnetized with a strength that mirrors the signal.

Later, when we want to play back the sound, the same tape is passed once again over the tape head, which now works in reverse, by electromagnetic induction. The varying magnetic field produced by the magnetized tape induces a corresponding voltage across the coil of the tape head, which can be amplified and fed to a loudspeaker to recreate the original sound. (In more elaborate tape recorders there are two different heads, one for recording and one for playing back, but the principle is still the same.)

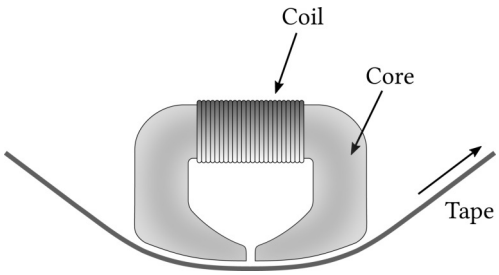


Figure 7.33: A tape recorder head. A tape recorder head consists of a coil wound around a specially shaped core. A sound signal is passed through the coil, creating a magnetic field that concentrates in the small gap at the bottom between the poles of the core. The moving tape passes over this gap and becomes magnetized.

There are a couple of points to notice about this process. First, as discussed in Section 7.2.7, the strength of the magnetic field produced by an electromagnet depends on the size of the current flowing in the coil, which means in turn that the amount of magnetization of the tape is roughly proportional to the sound signal. Upon playback, however, the voltage generated by the coil is proportional not to the magnetic field of the tape but to the rate at which the field is varying. Because the field varies faster for high frequency sounds than for low, this means that higher frequencies generate a larger voltage, other things being equal. As a result, the frequency balance of the sound gets distorted in the playback process, with too much treble and not enough bass.

In mathematical terms, the amplitude of the voltage, and hence the amplitude of the sound we ultimately produce, is proportional to frequency f and, bearing in mind that sound intensity is proportional to the square of amplitude (see Eq. (3.7)), this means that the intensity will increase by a factor of four, or 6 dB, when the frequency

goes up by one octave. In order for tape recordings to sound correct when we play them back, we have to correct for this issue. Luckily this is straightforward. As we saw in Section 7.2.4, a standard low-pass filter cuts out exactly 6 dB per octave above its cutoff frequency $f_c = 1/2\pi RC$. All we have to do to fix our tape recordings therefore is pass the output signal from the tape recorder through a low-pass filter with a low cutoff frequency—low enough to be below the bottom end of the hearing range at 20 Hz—so that all audible frequencies get filtered by 6 dB per octave. This process is known as *compensation*.

Second, the ability of tape to capture high-frequency sounds depends on the size of the gap between the poles of the tape head where the magnetic field is concentrated and on the speed at which the tape moves. If the tape is moving at velocity v and the distance between the poles is d, then it will take the tape an amount of time $\Delta t = d/v$ to cover this distance. Any change in the sound signal over a time shorter than this will get blurred out on the tape—the signals at two moments less than Δt apart will be recorded on overlapping portions of the tape and will not be distinct. This means that a sound has to have period greater than Δt to be recorded clearly, or equivalently that the frequency has to be less $1/\Delta t$. In other words, sounds are only captured faithfully up to a highest frequency of

$$f_{\text{max}} = \frac{1}{\Delta t} = \frac{v}{d}. \tag{7.74}$$

Rearranging this equation, we find that to achieve a particular f_{max} we need the speed of the tape to be greater than $v = df_{\text{max}}$. A typical size for the gap in a tape head is 10 micrometers, or 10^{-5} meters, so to record all audible frequencies up to 20 000 Hz the tape would need to move at a speed of

$$v = 10^{-5} \times 20\,000 = 0.2\,\text{m/s}, \tag{7.75}$$

or 20 centimeters per second. Standard reel-to-reel tape recorders use a speed of $7\frac{1}{2}$ inches per second, or about 19.1 cm/s, which is close enough to 20 cm/s that we should expect good reproduction across the full frequency range. Lower speeds are also sometimes used, at the expense of poorer high-frequency fidelity, and higher speeds are used in professional recording studios. Cassette recorders for consumer use employ a substantially lower speed of 4.75 cm/s, in order to fit the maximum amount of recording time onto a small tape, but they also use a tape head with a narrower gap to preserve high-frequency response. Nonetheless, cassette tapes have generally poor performance at the upper end of the frequency range, with the response dropping off significantly above about 15 000 Hz.

A cassette tape

7.4.2 Vinyl records

The very first sound recordings, in the late nineteenth century, used a groove cut in wax or foil to represent sound waves, and essentially the same technology is used

A long-playing record, or LP

in today's vinyl records. For much of the twentieth century the long-playing record, or LP, was the premier consumer medium for music reproduction, combining high sound quality with long running times and relatively sturdy construction. Although LPs are not competitive in any of these respects with today's digital recordings, they still enjoy popularity with collectors (for interesting reasons—see Section 8.4).

Sound on a vinyl record is captured in the shape of the groove that runs in a spiral around the surface of the record. The earliest recording media took the shape of a cylinder with grooves on the outside, but since the early twentieth century the preferred shape has been a flat disk, with grooves usually cut on both sides, so that one disk can hold two different recordings (such as two songs, or two halves of a longer musical work like a symphony). The groove wiggles from side to side in imitation of the waveform of the recorded sound, as shown in Fig. 7.34, and the sound is played back by placing a fine needle in the groove and then rotating the record, usually with an electric motor, making the needle vibrate back and forth in time with the wiggles in the groove. Just this on its own is enough to make a tiny sound—if you listen carefully you can hear sound coming from the needle of a record player as it vibrates. For proper listening, however, the sound must be amplified. To do this, the vibration of the needle is converted into an electrical signal using a mechanism similar to the dynamic microphone of Section 7.3.6: a coil is attached to the needle and a magnet placed nearby. The variation of the magnetic field through the coil as the needle vibrates generates a voltage by electromagnetic induction, which is amplified and then turned into sound using a loudspeaker.

The process of recording music onto vinyl records is laborious. The sound is normally captured on tape first, then turned into a record by cutting a groove into the surface of a metal master disk using a specialized lathe. The master disk is then used to cast a "negative" of the groove, with ridges sticking out where the grooves would normally go in. One such negative is cast for each side of the record. Then the actual LPs are manufactured by pressing molten plastic between the two negatives to create a disk with grooves on both sides. Although early records were made of wax, modern LPs are made of vinyl acetate, a form of plastic hard enough to resist damage, although it can still be scratched or broken by careless treatment, or warped by extreme heat. The possibility of damage is one of the reasons to prefer digital recordings over LPs.

Even when they are in perfect condition, however, LPs cannot match the sound quality of modern digital audio. A primary limitation is their *dynamic range*. The dynamic range of a sound recording, or of a musical performance generally, is the difference in decibels between the loudest and quietest parts. For example, a symphonic performance might range from 20 dB in moments of near silence to 90 dB when the entire orchestra is playing at full volume, giving a dynamic range of $90 - 20 = 70$ dB overall. In order to capture such a performance faithfully a recording medium needs

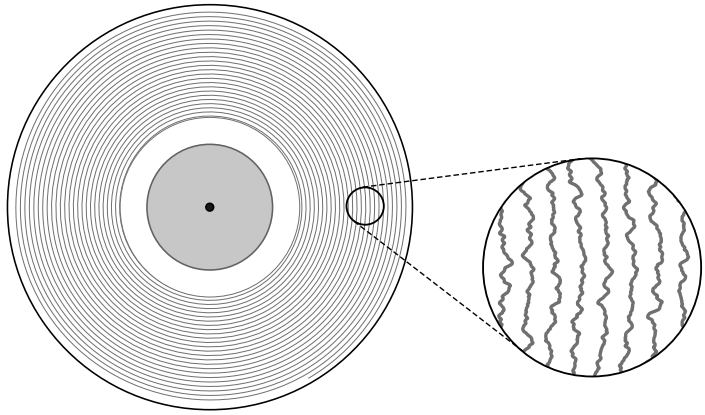


Figure 7.34: A vinyl record. Each side of a vinyl record is inscribed with a single groove that starts at the outside of the disk and spirals inward, making several hundred turns before ending near the center. If you examine the groove carefully (inset) you can see the wiggles that encode the waveform of the recorded sound.

to be able to record sounds over at least this range.

The quietest sound that can be recorded on a vinyl record corresponds to the smallest wiggle that can be reliably picked up by the needle and associated electronics in a record player. In practice this smallest wiggle is about 50 nanometers across. Similarly, the loudest sound corresponds to the largest wiggle, whose width is set by the need to fit a certain number of grooves on the record. For a typical LP that plays for 30 minutes a side at 33 revolutions per minute, the groove spirals around the record a total of $30 \times 33 = 990$ times in the space of about 10 cm, from its outermost point on the disk to its innermost. Thus the spacing of the grooves is around $0.1\,\text{m}/990 = 100$ micrometers, and the wiggles in the grooves are about half this size, or about 50 micrometers, to allow some room for error. Thus the ratio between the amplitude of the largest and smallest wiggles is

$$\frac{50 \text{ micrometers}}{50 \text{ nanometers}} = \frac{50 \times 10^{-6}\,\text{m}}{50 \times 10^{-9}\,\text{m}} = 1000. \tag{7.76}$$

The wiggles, as we have said, encode the shape of the sound waveform and hence their amplitude corresponds to the amplitude of the sound. Equation (7.76) tells us that the amplitude of the loudest sound that can be recorded is 1000 times that of the quietest. But recall that the *intensity* of a sound is proportional to the square of the amplitude (see Eq. (3.7)), so the loudest sound has $1000^2 = 1$ million times the intensity of the quietest. A factor of a million in intensity corresponds to a difference of 60 dB in sound intensity level (see Eq. (3.16)), so the dynamic range of our vinyl

record is about 60 dB—not enough to fully capture the 70 dB range of the orchestra in our example above.

It is possible to increase the spacing between the grooves on a record, also called the *pitch*, to allow larger wiggles and hence louder sounds to be recorded, at the expense of having fewer grooves overall and hence shorter recording time. It is also common to temporarily increase the pitch in louder portions of a piece of music to achieve a better dynamic range and then reduce it again in the quieter portions to save space. In this way the dynamic range can be pushed to about 70 dB, but that is the limit of the LP technology.

70 dB is enough to provide a good listening experience, but not enough to capture the true range of highs and lows in the most dramatic pieces of music. As we will see in Section 8.1.5, the dynamic range of a standard digital recording is 96 dB, significantly better than an LP, and enough to faithfully capture essentially any performance audible to the human ear.

EXAMPLE 7.5: PLAYING TIME OF A 7-INCH SINGLE

From the 1950s to the 1990s individual songs or other short recordings were commonly released on singles, vinyl records 7 inches in diameter that rotated at 45 revolutions per minute. The distance from the outer edge of the playing area to the inner edge on such a single is about 1 inch. If we want the record to have a dynamic range of 60 dB, how long in minutes can the recording be?

We have seen that the smallest wiggles in the groove of a record are about 50 nanometers across, which means the largest must be 50 micrometers in order to achieve a dynamic range of 60 dB (see Eq. (7.76) and the following discussion). Allowing 100 micrometers of separation between one turn of the groove and the next to give ourselves a safety margin, and noting that 1 inch is about 25 mm, we find that the maximum number of times the groove can wind around the record is

$$\frac{25 \times 10^{-3}}{100 \times 10^{-6}} = 250. \tag{7.77}$$

Given that the record spins at 45 revs per minute, the maximum playing time is then

$$\frac{250}{45} = 5.56 \text{ minutes}, \tag{7.78}$$

or about 5 minutes and 33 seconds.

This basic physical limit had a substantial effect on the development of 20th century popular music. The 7-inch single was for decades the primary format for the sale and promotion of popular songs, which meant songs had to be written to fit within about 5 minutes. It is possible to stretch the playing time a little longer, at the expense of lower dynamic range. *Hey Jude* by the Beatles, which was released as a 7-inch single, clocked in at 7 minutes and 11 seconds, which is around the maximum that can be achieved with the format. Don McLean's *American Pie* was 8 minutes and 42 seconds long and had to be split into two halves, one on each side of the record, when it was released as a single.

7.5 Loudspeakers

The final stage of the music reproduction process requires us to turn our electrical signals back into sound again so we can hear them, which is achieved using loudspeakers or headphones. Headphones are essentially just two loudspeakers mounted on either side of the listener's head, and the scientific principles of loudspeakers and headphones are basically the same. We focus here on loudspeakers.

Loudspeakers are the weak link in the music reproduction chain. Modern microphones can capture sound extremely faithfully, and today's recording and amplification technologies are virtually perfect. But even the best loudspeakers are problematic. Because of fundamental physical limitations, loudspeakers suffer both from uneven frequency response and haphazard directional output. In the following sections we will see why this is, and what can be done to improve the situation.

7.5.1 Dynamic loudspeakers

A dynamic loudspeaker

Loudspeakers come in a variety of different types but the most common by far, and the only one we will discuss at length, is the *dynamic loudspeaker*, which is the loudspeaker equivalent of a dynamic microphone: it uses a diaphragm, a coil, and a permanent magnet to turn electrical signals into sound.

Figure 7.35 shows the layout of a dynamic loudspeaker. The basic idea behind any speaker is the reverse of a microphone: we make a diaphragm vibrate and it pushes on the nearby air, causing it also to vibrate and hence making sound. The diaphragm of a dynamic loudspeaker, which is called a *cone* because it is cone-shaped rather than flat, is typically made of paper or sometimes thin plastic. It is fixed around its

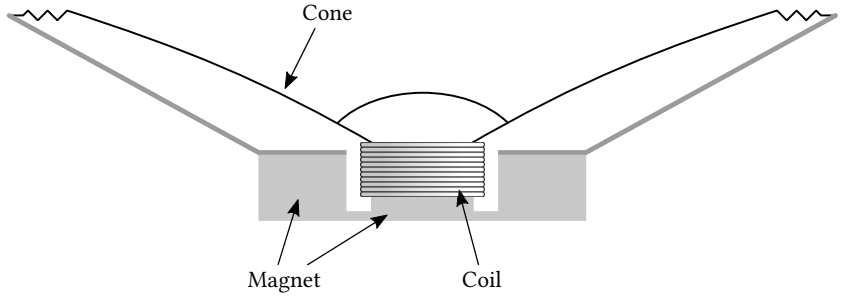


Figure 7.35: A dynamic loudspeaker. In a dynamic loudspeaker a diaphragm or cone, suspended at its edges, is driven back and forth by the attached coil. When electric current is passed through the coil it creates an electromagnet that is attracted to or repelled from the large permanent magnet mounted on the frame of the speaker. The motion of the cone causes the air to vibrate and produces sound.

outside edge to a sturdy frame and the edges are corrugated and flexible to allow the cone to move in and out. The rest of the cone is relatively rigid and maintains its shape as it moves. One can think of the cone as a circular piston, pushing the adjacent air and producing sound.

Mounted on the back of the cone in the center is a coil of wire similar to that in a dynamic microphone, and inside the coil is the pole-piece of a permanent magnet that is fixed to the frame of the loudspeaker and held stationary. When we pass current through the coil it creates an electromagnet with a strength that depends on the size of the current as described in Section 7.2.7, and this electromagnet is either attracted to or repelled from the magnet (depending on the direction of the magnetic field), which pushes or pulls the coil and moves the attached loudspeaker cone.

7.5.2 Sound production

Consider a loudspeaker driven by a pure sine-wave tone with frequency f. This causes the cone to move with a velocity u that also follows a sine wave $u(t) = u_0 \sin(2\pi f t)$, where the quantity u_0 represents the maximum velocity that the cone achieves. This motion in turn causes the air immediately against the front of the speaker cone to move with the same velocity.

Now recall that, as discussed in Sections 1.3 and 1.6.1, the sound pressure p in a sound wave is proportional to the velocity according to $p = zu$ (Eq. (1.2)), where the constant of proportionality z is the so-called acoustic impedance of air. This means that the pressure in this case will be $p(t) = zu_0 \sin(2\pi f t)$ and Eq. (3.7) then tells us that the instantaneous sound intensity is

$$\frac{p^2}{z} = z u_0^2 \sin^2(2\pi f t). \tag{7.79}$$

Then the average intensity is

$$I = z u_0^2 \left\langle \sin^2(2\pi f t) \right\rangle = \tfrac{1}{2} z u_0^2, \tag{7.80}$$

where we have made use of the fact that the average of $\sin^2(2\pi f t)$ is $\frac{1}{2}$.

Equation (7.80) tells us the intensity of the sound immediately in front of the speaker cone, i.e., the amount of sound energy per unit area per second. To get the total amount of energy P emitted by the speaker per second we multiply by the area A of the cone, which gives

$$P = AI = \tfrac{1}{2} A z u_0^2. \tag{7.81}$$

This is the power output by the speaker.

Because the power is proportional to cone area, we will get more sound from a large speaker than a small one. In simple terms, a large speaker pushes more air. Usually this means that we would like to make our speakers as large as possible, but

there are also disadvantages to large speakers as we will shortly see. Real speaker design, like so many other things, is a compromise between a number of competing factors.

Example 7.6: Sound output of a loudspeaker

Consider a typical small loudspeaker, circular with a diameter of 10 cm, playing a sine-wave tone at 500 Hz. About how much sound power will the speaker produce and how loud will the sound intensity be at 1 meter from the loudspeaker?

We can answer this question using Eq. (7.81), but first we need to know the maximum velocity u_0. Suppose the speaker cone is moving back and forth a distance of 0.1 mm, for a total of 0.2 mm traveled on each vibration (0.1 mm one way and 0.1 mm back again). It is doing this 500 times a second, meaning it is moving $500 \times 0.2 = 100$ mm or 0.1 meter each second, so its velocity is about 0.1 m/s. Meanwhile the radius of the circular speaker cone is $r = 5$ cm, so its area is $A = \pi r^2 = 0.0078\,\mathrm{m}^2$. The acoustic impedance of air is $z = 413\,\mathrm{Pa\,s/m}$ and, putting everything together and using Eq. (7.81), we get a sound power output of $P = 0.004$ watts or 4 milliwatts.

This is the total amount of energy put out by the loudspeaker per second. To calculate the intensity at 1 meter we use Eq. (6.52), which tells us that the intensity of direct sound from the loudspeaker at distance r is

$$I_{\text{dir}} = \frac{P}{4\pi r^2} = \frac{0.004}{4\pi \times 1^2} = 0.0003\,\mathrm{W/m}^2. \tag{7.82}$$

Converted to a sound intensity level using Eq. (3.11), this is equivalent to 85 dB. So even a small loudspeaker such as this can produce an impressive volume level. Moreover, this is probably an underestimate, because Eq. (7.82) is based on the assumption that the sound goes equally in all directions. As we will see in Section 7.5.6, in most cases loudspeakers project the bulk of their sound in the forward direction, making the sound level higher for a listener in front of the speaker.

7.5.3 Frequency response

Equation (7.81) tells us that the power output of a loudspeaker depends on the maximum velocity u_0 of the speaker cone. In Section 7.5.4 we investigate in detail how this velocity depends on the sound waveform, but even without delving into the mathematics we can understand roughly what will happen. First, the velocity of the cone is going to be smaller for low frequencies than for high, all other things being equal, because the cone is vibrating back and forth more slowly. This means that the power output of a speaker will be lower at low frequencies. However, the velocity of the cone is also lower at very high frequencies because the cone doesn't have enough time to respond to rapidly varying signals. At high frequencies the cone only just manages to start moving in response to an electrical input before that input reverses and it has to stop and move in the opposite direction. The cone's own inertia gets in its way and prevents it from responding.

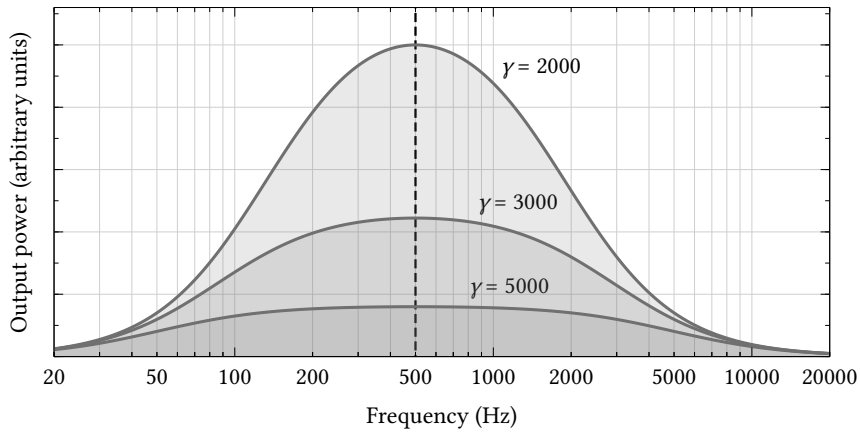


Figure 7.36: Output power of a loudspeaker. The output power of a loudspeaker, given by Eq. (7.91), peaks at the resonant frequency f_0 denoted by the dashed line, which is 500 Hz in this case. The height of the peak depends on the value of the damping parameter γ.

A speaker cabinet with separate speakers for different frequencies of sound.

The combined result of these effects is that the output power of a speaker looks like Fig. 7.36. It is poor both at low frequencies and at high frequencies, and best in the middle, with a peak at a characteristic frequency f_0, indicated by the dashed line in the figure. The characteristic frequency can be raised or lowered by changing details of the speaker design, but one cannot escape the fact that any real speaker only works well within a certain band of frequencies. If we want to cover the entire frequency range, therefore, we normally have to combine two or more speakers with different characteristic frequencies.

Even if we do this, however, we still have to contend with the fact that speakers used for low frequencies will have lower output power than similar speakers for high frequencies, because of the slower movement of the cone. Suppose we play a tone of frequency f through a speaker and suppose the distance the speaker cone moves as it vibrates—known as the *excursion*—is ξ. Then with each vibration in and out the cone is moving a total distance 2ξ and, since it is vibrating f times a second, the distance moved in one second, which is the average velocity of the cone, is $2\xi f$. The average velocity is not the same as the *maximum* velocity u_0 that appears in Eq. (7.81), but the two are proportional to one another, so the output power is

$$P = \tfrac{1}{2}Azu_0^2 \propto Az\xi^2 f^2. \tag{7.83}$$

Thus, for given area A and excursion ξ, the output power is proportional to f^2, meaning it will be higher for speakers that operate at high frequencies than for ones that

operate at low frequencies.[4]

We can compensate for the lower power output of a low-frequency speaker in two ways. Either we can make the speaker larger to increase the area A or we can increase the excursion ξ, which we do by driving the speaker with a stronger electrical signal. Most often we make the speaker larger, and this is one of the reasons why low-frequency speakers (woofers, subwoofers, bass amplifier speakers) are often large. Increasing the excursion is more difficult, since it can result in a distorted sound, but in recent years manufacturers have learned how to build high-excursion speakers that produce a clean bass sound with minimal distortion in a small package, leading to the current generation of small portable speakers, which sound remarkably good for their size.

At high frequencies Eq. (7.83) tells us that speakers should have ample power even with modest size and excursion, so high-frequency speakers can be small, and indeed work better if they are, since it reduces inertia which further improves high-frequency performance.

The bottom line is that no speaker can be good at all frequencies, but by varying their design one can make speakers that work well at different frequencies and then combine them to cover the whole range.

Advanced material

7.5.4 Movement of a speaker cone

The movement of a speaker cone results from a combination of forces acting on it. The mathematics is quite similar to that for the movement of the microphone diaphragm in Section 7.3.7. The driving force F_v felt by the cone comes from the magnetic coil and is proportional to the voltage $V(t)$ sent to the speaker:

$$F_v = bV(t). \tag{7.84}$$

The constant of proportionality b depends on the shape of the coil, the number of windings, the strength and placement of the permanent magnet in the speaker, and other design attributes.

There are also two other forces felt by the speaker. One comes from the suspension around the edges of the cone where it is attached to the speaker frame. The flexible edges of the paper cone act like a spring, allowing the cone to move in and out but pulling it back into place when the electrical input is turned off. To a good approximation the spring obeys the standard Hooke's law, which says that the restoring force F_r on the cone is proportional to its displacement $\xi(t)$ from its rest position, so

$$F_r = -k\xi(t), \tag{7.85}$$

where k is the spring constant and the minus sign indicates that the restoring force is in the opposite direction to the displacement.

The other force felt by the cone is a damping force, which acts to slow down the cone when it is moving. This force is proportional to the velocity of the cone $u = \mathrm{d}\xi/\mathrm{d}t$ and arises from a combination of several processes. There

[4] You might be concerned that Eq. (7.83) does not seem to square up with Fig. 7.36, since the equation appears to say that power always goes up with frequency while the figure shows it going up then down. The solution to this paradox is that the excursion ξ itself goes down at high frequencies, because of the inertia effects mentioned above, and this causes the value in Eq. (7.83) to fall off when the frequency gets too high.

is friction in the paper of the cone itself, as well as air resistance from the air around it, and there is also an electrical phenomenon called *back electromotive force* or back EMF, in which the motion of the coil of the loudspeaker induces a current in the coil, similar to that in a dynamic microphone, thereby turning some of the cone's kinetic energy into electrical energy and so reducing the kinetic energy. Together these effects combine to produce a damping force that can be written as

$$F_d = -\gamma \frac{\mathrm{d}\xi}{\mathrm{d}t}, \tag{7.86}$$

where γ measures the magnitude of the damping and the minus sign represents the fact that the damping force is in the opposite direction to the velocity.

Combining the three forces acting on the cone and applying Newton's second law, we get $F_v + F_r + F_d = ma$, where m is the mass and a the acceleration of the moving cone and its attached coil. Using Eqs. (7.84), (7.85), and (7.86) and writing $a = \mathrm{d}^2\xi/\mathrm{d}t^2$, we then find that

$$m\frac{\mathrm{d}^2\xi}{\mathrm{d}t^2} + \gamma\frac{\mathrm{d}\xi}{\mathrm{d}t} + k\xi = bV(t). \tag{7.87}$$

Now suppose that the input to the speaker is a sine wave with frequency f, which we will represent as the real part of a complex voltage $V(t) = V_0\mathrm{e}^{\mathrm{i}\omega t}$ with $\omega = 2\pi f$. Then the steady-state solution to Eq. (7.87) takes the form $\xi(t) = \xi_0\mathrm{e}^{\mathrm{i}\omega t}$ and, substituting into the equation and cancelling some factors, we find that $-\omega^2 m\xi_0 + \mathrm{i}\omega\gamma\xi_0 + k\xi_0 = bV_0$, or

$$\xi_0 = \frac{bV_0}{k - \omega^2 m + \mathrm{i}\omega\gamma}. \tag{7.88}$$

The velocity $u(t)$ of the cone is equal to the derivative of the displacement

$$u(t) = \frac{\mathrm{d}\xi}{\mathrm{d}t} = \mathrm{i}\omega\xi_0\mathrm{e}^{\mathrm{i}\omega t}, \tag{7.89}$$

which means the velocity also follows a sine wave and therefore produces a sine-wave sound pressure, as we would hope. The amplitude u_0 of the velocity is given by the complex magnitude of (7.89), which is

$$u_0 = \left|\omega\xi_0\right| = \frac{\omega bV_0}{\sqrt{(k - \omega^2 m)^2 + \omega^2\gamma^2}}. \tag{7.90}$$

Substituting into Eq. (7.81), the output power of the speaker is then given by

$$P = \frac{\frac{1}{2}\omega^2 Azb^2V_0^2}{(k - \omega^2 m)^2 + \omega^2\gamma^2}. \tag{7.91}$$

Figure 7.36 shows how this power varies with frequency for typical choices of the parameters and, as we can see, the output is low for both high and low frequencies and peaks in the middle. There is a characteristic resonant frequency f_0 of the speaker at which it produces its greatest output, which we can calculate by differentiating Eq. (7.91) with respect to ω and setting the result to zero. After some algebra, we find that

$$f_0 = \frac{1}{2\pi}\sqrt{\frac{k}{m}}. \tag{7.92}$$

For frequencies around f_0 the speaker will work well, but if we get too far away from this frequency the speaker will have poor output power. By varying the spring constant k of the speaker cone and the mass m, we can make the resonant frequency higher or lower and hence create speakers that are suited for high- or low-frequency use.

The height of the peak in Fig. 7.36 depends on the damping parameter γ, going down as the damping is increased. Curves for three different choices of γ are shown in the figure. Having a lower peak is a good thing in some ways—it makes the speaker's output more uniform over the frequency range—but it also reduces the overall power level, so there is a trade-off between having a speaker with a more uniform response and having one with good output power.

7.5.5 Enclosures

A practical problem with loudspeakers is that sound emanates from both the front and the back of the speaker cone, but the waveform coming from the back is upside-down relative to the one coming from the front. When the cone pushes on the air at the front it pulls on the air at the back and *vice versa*. This gives rise to phase

cancellation as discussed in Section 4.5. Two sound waves, identical except that one is inverted relative to the other, will cancel each other out and produce no sound. The cancellation of sound from front and back of a speaker is not perfect: assuming the listener is in front of the speaker the sound from the front will normally be louder than the sound from the back, and there is typically also a time delay for the sound to reach the listener from the back which means that the two waveforms are not exact opposites of one another. Nonetheless, a significant amount of cancellation can occur, reducing the apparent volume of the sound. In technical language, we say the loudspeaker is a "dipole radiator" of sound, meaning it is a combination of high pressure on one side of the cone and low pressure on the other.

The solution to this problem is a simple one: we mount the loudspeaker in the wall of a closed box or enclosure. The front of the speaker faces outward from the box and projects sound toward the listener but the back is inside where sound cannot escape. The enclosures that loudspeakers come in are not just for looks or convenience—they play an important role in the sound.

As discussed in Section 7.5.3, we often use multiple separate loudspeakers of different sizes to cover the whole frequency range, in which case all of the speakers can be mounted in the same enclosure, which is the typical design for consumer loudspeakers.

7.5.6 Directionality

Loudspeakers are directional, meaning that they project sound most strongly to the front and less to the sides. This can be detrimental to sound quality for listeners who are not directly in front of the speaker. Moreover, the effect is more pronounced for high frequencies than for low, so that a listener situated to the side may hear a distorted balance of frequencies with too much bass and not enough treble, resulting in a muffled sound. To see why this happens, consider Fig. 7.37, which shows a sketch of the sound waves emanating from a loudspeaker.

An exception to the issues discussed here is the speakers in headphones, also called "drivers." Because they are always pointed directly at our ears, off-axis sound quality is normally not an issue with headphone drivers.

Suppose once again that our loudspeaker is playing a sine-wave tone with frequency f and consider sound traveling not directly away from the front of the speaker but off-axis, at an angle θ as shown. For an off-axis listener hearing this sound, the sound coming from different parts of the speaker cone has to travel different distances to reach their ears. If the speaker has diameter d then sound from the left edge of the speaker in the figure has to travel a distance $d \sin\theta$ further than sound from the right edge. This means that sound waves from different parts of the speaker are not perfectly in phase with one another and can cancel each other out, either partially or completely, decreasing the apparent volume of the sound.

In the situation shown in the figure, for instance, the two highlighted waves, the first and fourth ones from the left, have exactly opposite phases. Everywhere that one wave goes up the other one goes down and *vice versa*. (Use the dashed lines to

guide your eye.) Thus these two waves cancel out exactly and produce no sound. The same is also true of the second and fifth waves, and the third and sixth—each wave has a partner that exactly cancels it. The result is that in this case *all* of the sound traveling at angle θ cancels out and a listener at this angle will hear nothing at all.

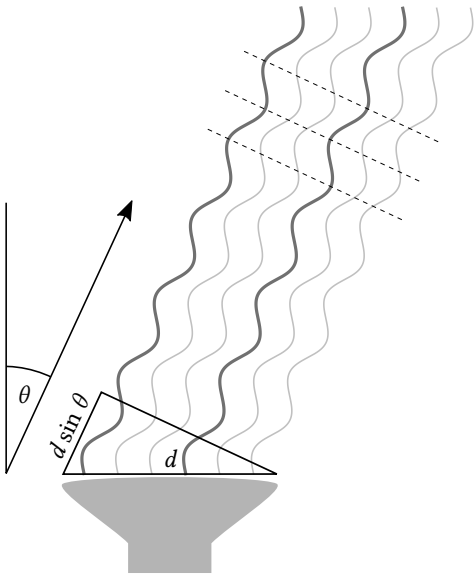


Figure 7.37: Sound leaving a speaker at an angle. Sound waves emanating from a loudspeaker at an angle θ from the central axis can be partially or completely out of phase with one another and so cancel out. The two highlighted waves, for example, have exactly opposite phases and hence cancel completely.

Mathematically speaking, this total cancellation happens when the phase of the waves goes through one complete cycle from one side of the speaker to the other. Look again at Fig. 7.37 and you will see that from the left side of the speaker to the right the phase slowly varies through an entire cycle until it comes back into phase with itself. This means that the extra distance $d \sin\theta$ traveled by the leftmost wave in the picture relative to the rightmost one is exactly one wavelength λ of the sound, so $d \sin\theta = \lambda$, or equivalently

$$\sin\theta = \frac{\lambda}{d}. \tag{7.93}$$

This equation tells us the critical angle θ at which the sound will completely cancel out. From the value of $\sin\theta$ we can calculate θ itself by taking the arcsine.

If the listener is off axis by less than this critical angle then the sound will not completely cancel and they will hear the tone from the loudspeaker, although there will be partial cancellation as they approach the cutoff, so the sound will get quieter. In practice this means that there is a region, or cone, in front of the speaker, of angular width θ given by Eq. (7.93), and for clear sound we want to be inside this cone and not too near its edges. Figure 7.38 shows a computer calculation of the actual sound level from a loudspeaker, with the edges of the cone marked by dashed lines.

In practice, we would like our speakers to work no matter where the listener is situated, which means we want to make the cone of projection of the speaker at least 90° wide, so that it occupies the entire space in front of the speaker. Putting $\theta = 90°$ in Eq. (7.93) and noting that $\sin 90° = 1$, this tells us that to be able to hear the sound from all directions we need λ/d to be significantly larger than 1, which means the diameter d should be significantly less than the wavelength λ. This places a limit on the size that we can make a loudspeaker if we want it to sound good in practice. (This is in addition to considerations about the inertia of the cone discussed in Section 7.5.3.)

From Eq. (2.4) we know that wavelength is related to frequency by $\lambda = c/f$ (where c is the speed of sound), so the size limit on a loudspeaker depends on the fre-

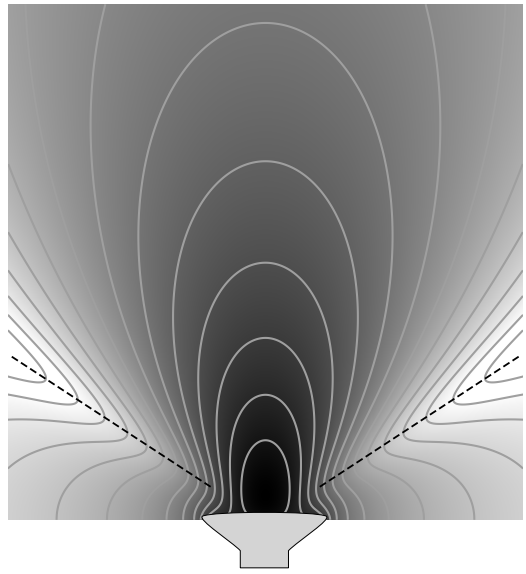


Figure 7.38: The sound projected by a speaker. This plot shows a computer calculation of the sound pressure level produced by a circular speaker 25 cm in diameter (shown at the bottom) playing a sine-wave tone at 2000 Hz. Darker shades indicate higher SPL and lighter shades lower. The dashed lines are the theoretical edges of the cone of projection, although most of the sound is projected well inside this cone (and a small amount outside it).

quency of the sound we want to reproduce. Higher frequencies correspond to shorter wavelengths and hence require smaller speakers. This is another reason why high-frequency speakers (also called *tweeters*) are generally smaller than low-frequency ones (*woofers*). They have to be smaller to project the sound widely. Smaller speaker area does mean less output power, as discussed in Section 7.5.2, but this is offset by the fact that the cone moves faster at higher frequencies and produces higher sound pressure, so usually the smaller size is not a problem.

The calculation here is not exactly correct because it does not take account of the fact that the speaker is circular. A full calculation, which you can find in Section 7.5.8, gives a slightly different answer, that in fact a loudspeaker should be smaller than 1.22λ in order to project sound in all directions. In practice, the difference is minor. It's fine to use the full formula, but the rough rule of thumb that d should be less than the wavelength is easier to remember and is good enough for most purposes.

Example 7.7: Sizes of loudspeakers

Consider sound in the bass frequency range, at say 250 Hz. How large can a loudspeaker be if it is to work well in this frequency range?

The corresponding wavelength is $\lambda = 343/250 = 1.4$ meters and the speaker should be significantly smaller than $1.22\lambda = 1.7$ meters, so a safe choice in this case would be to make the speaker well under meter in diameter. In practice, the largest size one sees for bass speakers is about half a meter.

And what about the other end of the scale, at high frequencies? At the top of the human hearing range around 17 000 Hz, the wavelength is $\lambda = 343/17\,000 = 2$ cm and the speaker should be smaller than $1.22\lambda = 2.5$ cm. In practice, tweeters are right around this size and not much smaller, which means we are sacrificing some off-axis performance at the high end of the frequency range. The musical contribution of the very highest frequencies, above say 15 000 Hz, is only modest however—and for many people actually inaudible—so this is perhaps a reasonable compromise.

7.5.7 Design of loudspeaker systems

Putting together what we have learned, we can now say something about the design of complete speaker systems. There are three main issues at play. First, as we have seen, the mechanics of the speaker cone favors a band of frequencies centered on the characteristic frequency f_0, whose value can be varied to create a speaker that favors low-, mid-, or high-frequency bands, but not all of them at once. If we want to cover the entire frequency range we normally have to combine two or more speakers tuned to different bands.

Second, speakers generally produce stronger sound at high frequencies than low, all other things being equal, because the speaker cone moves faster at high frequencies and generates higher sound pressure. We can compensate for the lower output power of low-frequency speakers either by increasing the area of the speaker or by increasing its excursion—the range of movement of the cone—which we do by driving the speaker with a stronger electrical signal. Historically, speaker designers usually took for the former approach and built larger speakers, since increasing the excursion posed some technical challenges. These difficulties have however mostly been overcome in recent years, leading to the creation of small, high-excursion speakers that offer good bass output despite their diminutive size.

The third issue is directionality. Speakers tend to project sound most effectively to the front and less to the sides, and this effect is more pronounced for large speakers and at high frequencies. In order to project well over a wide angle, the diameter of a speaker should be less than the wavelength of the sound, which means in practice that speakers intended for high frequencies must be small—a few centimeters across at most. Speakers intended for low frequencies on the other hand can be large without compromising directionality.

Combining these observations we can now gain an understanding of how to build an effective speaker system. First, there are some situations where we need to reproduce only a limited frequency range. Amplifiers for bass guitar, for instance, typically aim to reproduce sounds only up to about 1000 Hz, which can be done with a single speaker, usually large since directionality is not an issue and higher output power is desirable—30 or 40 cm speakers are not unusual. In the more common case where we want to reproduce the entire audible frequency range, we combine multiple speakers, small tweeters for the high frequencies and woofers and subwoofers for the low frequencies which may or may not be larger. Traditional speaker designs, including most high-quality consumer models produced until the 1990s, used large low-frequency speakers for their superior output power, resulting in good sound but bulky construction. Commercial amplification and public address systems, for which space and weight are not an issue, still use this approach. Recent years, however, have seen the rise of small portable speakers for home use that produce surprisingly good sound even in the bass range by using larger excursions instead of larger size.

Finally, in order to prevent sound from the back of a speaker cancelling sound from the front, it is important as we have seen that all speakers be mounted in a box or enclosure, the front of the speaker facing outward while the back is sealed inside the box to prevent sound escaping.

Example 7.8: Designing a two-speaker system

Suppose we want to build a full-frequency-range loudspeaker system with two individual speakers in a single enclosure, a woofer and a tweeter. The woofer is to have a characteristic frequency of $f_L = 400$ Hz and a maximum frequency of 2000 Hz, while the tweeter is to have characteristic frequency $f_H = 7000$ Hz and span the range from 2000 Hz to 18 000 Hz. Also the tweeter will be limited to a maximum excursion of 1 mm, which is typical for high-frequency speakers. Suggest suitable sizes for the two speakers that will give good directionality for both. What will the maximum excursion of the woofer need to be if we want the maximum output power to be the same for both speakers at their characteristic frequencies?

A number of factors come into play here. The maximum size of the speakers is dictated by directionality, as discussed in Section 7.5.6. For best performance we need them to be smaller than 1.22 times the wavelengths of their highest frequencies, which are

$$\lambda = \frac{c}{f} = \frac{343}{2000} = 17\,\text{cm} \tag{7.94}$$

for the woofer and

$$\lambda = \frac{343}{18\,000} = 1.9\,\text{cm} \tag{7.95}$$

for the tweeter. This gives us upper limits of 21 cm and 2.3 cm on our speakers, so let us say the diameters are 15 cm and 1.5 cm, to give us some margin for error.

At the same time, Eq. (7.83) tells us that output power at frequency f is proportional to $A\xi^2 f^2$, where A is the area and ξ is the excursion. Writing the area in terms of the

diameter d as $A = \frac{1}{4}\pi d^2$, we then have output power $P \propto d^2 f^2 \xi^2$. If we want our two speakers to have the same power we then need

$$d_L^2 f_L^2 \xi_L^2 = d_H^2 f_H^2 \xi_H^2. \tag{7.96}$$

Rearranging for the excursion ξ_L of the low-frequency speaker, we get

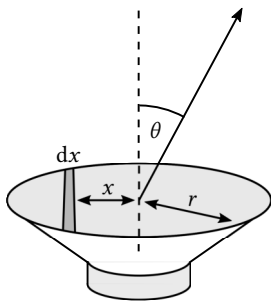
$$\xi_L = \frac{d_H f_H}{d_L f_L} \xi_H = \frac{1.5 \times 7000}{15 \times 400} \times 1\,\text{mm} = 1.75\,\text{mm}. \tag{7.97}$$

So the woofer needs to have a maximum excursion of about 2 mm, twice that of the tweeter.

ADVANCED MATERIAL

7.5.8 DIRECTIONALITY OF A CIRCULAR SPEAKER

Calculation of the directionality of loudspeakers is made more complicated by the fact that most speakers are circular. Consider a slice of a circular speaker cone a distance x from the center line and of width dx, thus:



For a listener situated at an angle θ off axis, the extra distance traveled to the listener by sound from this slice, compared to sound coming from the center line, is $x \sin\theta$. We ignore the slight variation in θ for different x due to the changing direction to the listener, as well as the small difference in distance from different points along the slice. This approximation, known as the "far field," is a good one as long as the listener is sufficiently far away, which in practice means that the distance to the listener should be much greater than the speaker diameter, and this is true in most settings. (We briefly discuss two situations where it is not at the end of this section.)

Given that sound travels at speed c, it will take a time $(x/c)\sin\theta$ to travel the extra distance. If the loudspeaker is producing a sine-wave tone with sound pressure proportional to $\sin(2\pi f t)$, then the delayed sound from our slice will have pressure proportional to $\sin[2\pi f(t - (x/c)\sin\theta)]$, which can be conveniently represented as the imaginary part of the complex expression

$$e^{i2\pi f(t-(x/c)\sin\theta)} = e^{i\omega t} e^{-ikx\sin\theta}, \tag{7.98}$$

where $\omega = 2\pi f$ and $k = 2\pi f/c = 2\pi/\lambda$.

Pythagoras' theorem tells us that the length of the slice is $2\sqrt{r^2 - x^2}$ where r is the radius of the speaker and, assuming an equal amount of sound from each part of the slice, we can add up all of the incoming sound to calculate the sound from the whole speaker. (There will be a decay of intensity with distance to the listener, due to the inverse square law, as described in Section 3.5, but in the far field this will be approximately the same for sound from all parts of the speaker, so we can ignore it.) The total sound pressure felt by the listener will then be

$$\begin{aligned} p(t) &\propto e^{i\omega t} \int_{-r}^{r} e^{-ikx\sin\theta} \sqrt{r^2 - x^2}\, dx \\ &= \pi e^{i\omega t} r \frac{J_1(kr\sin\theta)}{k\sin\theta}, \end{aligned} \tag{7.99}$$

where $J_1(x)$ is a Bessel function, a special function that appears commonly in calculations with circular symmetry such as this one. The intensity I of the sound heard by the listener is proportional to the magnitude squared of the sound pressure, meaning that

$$I = A\pi^2 r^2 \left(\frac{J_1(kr\sin\theta)}{k\sin\theta}\right)^2, \tag{7.100}$$

where A is a proportionality constant. Making use of the fact that $J_1(x)/x \to \frac{1}{2}$ as $x \to 0$, the value of this intensity

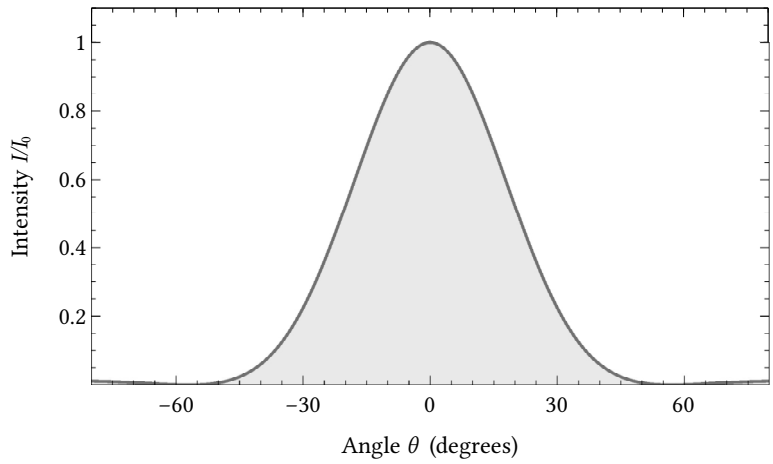


Figure 7.39: Angular output of a circular loudspeaker. The intensity of the sound from a circular loudspeaker 25 cm in diameter playing a 2000 Hz sine-wave tone, as heard by a listener at an angle θ off-axis and measured as a fraction of the on-axis intensity I_0.

on-axis at $\theta = 0$ is $I_0 = \frac{1}{4}A\pi^2 r^4$, and hence

$$I = 4I_0\left(\frac{J_1(kr\sin\theta)}{kr\sin\theta}\right)^2. \tag{7.101}$$

Figure 7.39 shows a plot of this quantity for the case of a 25 cm loudspeaker playing a tone at 2000 Hz. As we can see, the loudspeaker only produces significant intensity within an angle of about 30°. This would not be a satisfactory speaker in practice, at least for sound at this frequency. The sound would be too attenuated for off-axis listeners.

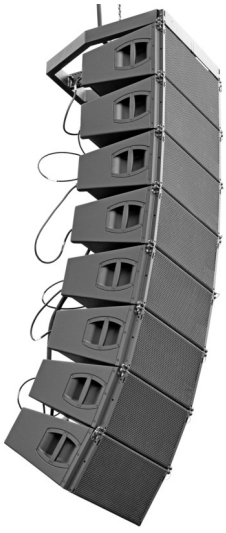
The first zero of the Bessel function $J_1(x)$ occurs at $x = 3.8317$, which means Eq. (7.101) falls to zero when $kr\sin\theta = 3.8317$. Rearranging and setting $k = 2\pi/\lambda$ and $r = \frac{1}{2}d$ where d is the diameter of the speaker, we find that the angular width θ of the cone of projection of the speaker satisfies

$$\sin\theta = 1.220\,\frac{\lambda}{d}. \tag{7.102}$$

For the case in Fig. 7.39, for instance, with $d = 0.25$ m and $f = 2000$ Hz, this gives $\theta = 56.8°$, and we need to be well within this angle for good sound from the speaker. (Figure 7.39 suggests that about 30° or less would be desirable, as we have said.)

As discussed in Section 7.5.6, we would ideally like the sound projected by our loudspeaker to extend all the way to $\theta = 90°$, which means $\sin\theta = 1$. Making this choice in the equation above tells us that we would need the diameter of the speaker to be significantly less than 1.22λ. This is close to the answer we got in Section 7.5.6, where our rough calculation indicated that the speaker should be smaller than λ. The calculation here is more correct, taking proper account of the circular shape, but the simple rule of thumb that the speaker should be significantly smaller than the wavelength is good enough for most purposes.

The far-field approximation used here is a good one for most loudspeakers. Consumer speakers in people's homes, musical instrument amplifiers, and public address systems are all normally heard by listeners at some distance from the speaker. Two situations where this is not the case are headphone speakers, which are mounted close to the listener's ears, and studio monitor speakers used by mixing engineers, which are normally placed on the desk immediately in front of the engineer in order to minimize the effect of room acoustics. Calculations for these speakers are different and the speakers themselves must be designed to allow for the fact that they will be used in a different way. Studio monitors are sometimes labeled as "near-field" monitors, to emphasize the fact that they are designed to be heard close up.



A phased array consisting of eight speaker enclosures stacked vertically.

7.5.9 Phased arrays

Directionality is not always a bad thing in loudspeakers. Particularly for large public address systems, such as are used at live music concerts, being able to project sound in a particular direction can be good. For instance, in an open-air performance situation there is little point projecting sound from the stage up into the sky when all the listeners are on the ground. Large-diameter loudspeakers can be useful in such situations for their directionality, but a more common approach is to use a *phased array*.

A phased array, also called a line array in the music industry, is a line of speakers, usually stacked vertically, all playing the same sound. In effect the phased array acts as a single tall narrow speaker. Because of its great height—often several meters—the array projects into a narrow angle vertically, but because its width is comparatively small it projects into a wide angle horizontally. The result is that the phased array projects a horizontal sheet of sound that is perfect for broadcasting to an audience on the ground or floor of a performance space. All of the power generated by the speakers is concentrated on the audience and none is wasted on the sky.

Example 7.9: Phased arrays

A phased array speaker is 3 meters tall and 50 cm wide. Into approximately what angle does it project sound horizontally and vertically at 1000 Hz?

For the purposes of this question we can treat the phased array as a single rectangular speaker, which therefore obeys Eq. (7.93) on page 266. The wavelength of sound at 1000 Hz is $\lambda = c/f = 343/1000 = 34$ cm and, setting the diameter to 50 cm, the horizontal angle is $\sin\theta = 34/50 = 0.69$. Taking the arcsine then gives us $\theta = 43°$, which is an acceptable projection angle for a concert stage. On the other hand the vertical angle has $\sin\theta = 34/300 = 0.115$, which gives $\theta = 6.6°$, a much narrower angle. So this speaker will project a sheet of sound that is very narrow vertically but covers a wide swath of the audience—exactly what a phased array is designed to do.

7.6 Stereo recording and surround sound

When we hear a live musical performance, sound often comes at us from many different directions. The musicians in a band or orchestra may be scattered about a stage, for instance, and the ear makes use of the different directions of sounds to tease apart the music and interpret what we are hearing, as discussed in Section 5.5. Sound may also be reflected off the walls, floor, and ceiling of a room as discussed in Section 6.7, coming at us from all directions and creating the sensation known as "envelopment," which contributes substantially to the impact of live sound.

Conversely, recorded sound, played over a single loudspeaker—so-called *monophonic* sound—comes at us, by and large, from just one direction. Sound may bounce

off the walls to create a certain degree of envelopment, but most recorded music is heard in acoustically dry environments such as people's houses and cars, which reflect little sound. The result is that sound from a single speaker lacks some of the elements that lend live music both its grandeur and its clarity. *Stereophonic sound* (or stereo for short) and its sibling *surround sound* are an attempt to recreate these lost elements, at least in part, by simultaneously playing sound over two or more loudspeakers spread out around a room.

Developed from the 1930s onward and finding widespread use starting in the 1960s, stereo sound is played over two loudspeakers. Two slightly different recordings of the same music are made and played back over each of the speakers, creating a *stereo image*—the acoustic impression of sound coming from a wide spatial field. Ideally the listener sits (or stands) with the speakers in front of them and separated by a certain angle—60° is often recommended. The sound of an instrument played only on the left speaker will appear to be coming directly from that speaker, i.e., from 30° to the left, or the 11 o'clock position. Similarly sound played only over the right speaker will appear to be coming from 30° to the right, or the 1 o'clock position. And sound that comes equally from both will appear to be coming from a central position, 12 o'clock, directly in front of the listener. By careful adjustment of the relative signal strength in the two speakers, one can make sound appear to emerge from any direction in the 60° stereo field, and by giving different instruments a different balance of left and right we can make them each appear to be located in their own spatial position and hence create the impression of a band or orchestra arrayed in space in front of the listener.

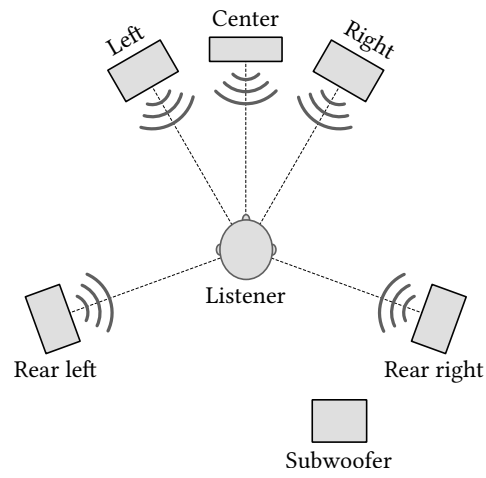


Figure 7.40: A typical speaker arrangement for surround sound. In a typical six-speaker surround sound arrangement there are left and right speakers in front of the listener at about 30° to either side, a center speaker, and two rear speakers about 110° to either side. The optional subwoofer can be placed almost anywhere, since the ear is insensitive to the direction of the very low-frequency sounds it produces.

Stereo can capture the sounds of different instruments in different locations, but it cannot capture the envelopment produced by reverberated sound reaching the listener from all directions in a concert hall or performance space. For that we need surround sound, an extension of stereophonic sound that makes use of more than two speakers, some placed behind the listener, so that he or she is surrounded by sound on all sides. The most common version of this idea uses six speakers as shown in Fig. 7.40. There are left and right stereo speakers as normal, a center speaker in between those two, and another stereo pair behind the listener to the left and right. The sixth speaker, which is optional, is a subwoofer, devoted to the production of the lowest sound frequencies. Because

the ear is poor at telling which direction low-frequency sound comes from (see Section 5.5.2), the subwoofer can be placed almost anywhere within hearing range. Its location is not important for the surround-sound effect.

You might wonder what the point is of the center speaker if we already have left and right speakers on either side of it. We can produce the effect of sound coming from the center of the stereo field by playing it equally through the left and right speakers, so why do we need the center speaker? In fact, for musical applications we do not really need it. Surround-sound systems without a center speaker are fine for listening to music. The main role of the center speaker is in audiovisual applications, such as watching TV or movies. For the sake of realism, it is important when watching a movie that sound from events seen on the screen, such as actors talking, appear to *come* from the screen, i.e., from the center of the stereo field. This can, as we say, be achieved by playing the sound equally over the left and right speakers, but this only works if the listener is situated in the perfect listening position exactly in front of the speakers. As we saw in Section 5.5, perception of the direction of sound is based on differences in loudness and timing between a listener's two ears, and both loudness and timing can be affected when the listener is off-center. A listener sitting to the left of center, for instance, will hear sound from the left speaker that is both louder than the right and arrives sooner, both of which will make the sound seem to come from the left side of the stereo field and not the center. For a movie watcher this can have the disorienting effect of making actors' voices appear to be coming from a different place from where their mouths are. The center speaker rectifies this problem by providing a direct sound source that is always in the same place as the screen. In surround-sound film and video most dialog and other on-screen sound is played through the center speaker. As an additional benefit, this approach also allows one to separately adjust the volume of the dialog relative to other sounds by changing the volume of the center speaker.

In movies, surround sound is used to generate special effects such as footsteps coming from behind the audience or a plane flying overhead. When used for music, however, its primary virtue lies in its ability to create envelopment. By playing direct sound from the musicians over the front speakers and reverberated sound, as if reflected off the walls, over the rear speakers, one can create an approximate facsimile of the experience of hearing live music in a large concert venue.

This at least is the ideal. The practice, unfortunately, often falls short, for a variety of reasons. First, music has to be recorded in stereo or surround sound to get a stereo or surround effect. Since the late 1960s, most music has been recorded in stereo, so stereo recordings are now very common, but surround-sound recordings are not. Music for film soundtracks, intended to be heard over the multispeaker systems in movie theaters, has long been recorded with surround sound in mind, but surround-sound music for home listeners has only started to become widely available in the

streaming music era since the mid 2010s.

Second, to get the full effect of stereo or surround sound, a listener has to be positioned exactly in the middle of the speakers, equidistant from left and right channels. While this may be possible when watching a movie, it is not common when listening to music. Busy people listen to music while doing other things and are rarely sitting still in the ideal position. And a lot of music is heard in the car, where it may not even be physically possible to sit half way between the speakers. Moreover, although stereo hifi systems were the dominant form of consumer audio equipment until the early 21st century, the increasing popularity of smaller portable and wireless speakers has made true stereo sound rarer in the home in recent years. Some portable speakers do make an attempt to project stereo sound using clever internal machinery, but the effect is inferior to a true stereo speaker pair with wide separation and the resulting sound is essentially only monophonic, especially when heard from some distance away.

A more subtle issue arises when one listens to music over headphones or earphones. In that case one normally hears the sound intended for the left stereo speaker in the left ear and the right speaker in the right ear. This, however, produces a completely different sensation from listening over loudspeakers. When you hear sound from a loudspeaker, that sound will, to some extent at least, reach both of your ears. Even if the speaker is located to your left, for instance, you will hear some sound from it in your right ear. This allows you to tell what direction the sound is coming from—as discussed in Section 5.5, your brain uses differences in loudness and timing between your ears to determine direction. When you listen to recordings over headphones, on the other hand, your brain can no longer do this, since the sound for each speaker is now heard by only one ear. The result is that, while your two ears do hear different things when listening on headphones, the stereo image is severely compromised. It is possible to rectify the situation using special "binaural" recordings designed to be heard on headphones, that deliberately introduce differences in loudness and timing between the two ears to mimic the effect of sounds coming from different directions and trick the brain into perceiving a stereo field. Indeed binaural recordings arguably give the best illusion of space and envelopment in music, better than standard stereo recordings played over loudspeakers, because of the ability to control very tightly what each ear hears. Unfortunately, true binaural recordings are rare, although in recent years a number of manufacturers have introduced hardware that can mimic binaural audio electronically and create an illusion of surround sound over headphones or earphones.

Technologies were developed in the first half of the 20th century to allow stereo and surround sound to be captured on magnetic tape and vinyl records. On tape the principle is simple: two or more separate "tracks" of sound are captured side by side on the same tape by magnetizing different strips of the tape with different recording

heads. Playing back the tape similarly involves using two or more heads to recover the separate recordings, along with two or more amplifiers to amplify them and two or more speakers to produce the sound.

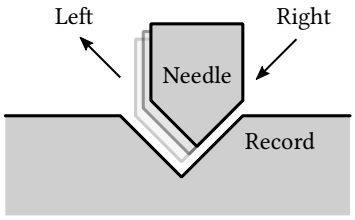


Figure 7.41: The groove in a stereo LP. Left and right channels of a stereo recording are encoded in the groove of a record as movements of the groove diagonally, up and to the left for the left channel, down and to the left for the right channel.

For vinyl records things are a little more complicated. Instead of using a groove that moves from side to side only, the groove on a stereo record is also allowed to move up and down. The two channels of the stereo are encoded in diagonal movements of the groove up-and-to-the-left or down-and-to-the-left, as shown in Fig. 7.41. The record is played back with a needle that picks up motion along the two diagonal directions separately, creating two separate signals that are amplified for playback over separate speakers. This design allows the record to hold two independent channels while still ensuring that a monophonic record, in which the groove moves from side to side only, will play correctly on a stereo record player, with sound coming from both channels.

With digital recordings, there are various ways to capture stereo sound, the simplest being just to store digital versions of the right and left stereo channels separately. A similar approach can be used for surround sound, allowing us to store six separate channels or even more. We discuss digital sound in more detail in Chapter 8.

Chapter summary:

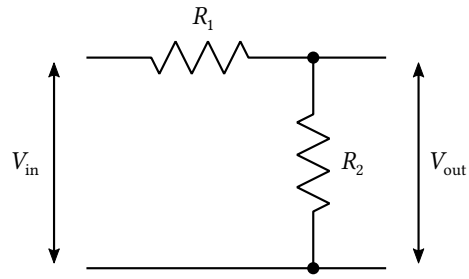
- The recording and reproduction of sound is achieved by turning sound into analogous **electrical signals** and then back into sound again.
- The electrical signals can be manipulated in a variety of useful ways using electronic circuits, for instance to increase or decrease the volume or filter out certain frequencies from the sound.
- **Microphones** capture sound and turn it into electrical signals. A microphone consists of a diaphragm, usually made of mylar, that senses pressure changes, a capsule to contain the diaphragm, and some means of detecting the movement of the diaphragm.
- Microphones come in two main varieties. **Pressure microphones** directly measure pressure on the diaphragm using a capsule closed at the back, and are omnidirectional. **Pressure-gradient microphones** measure the pressure difference between the two sides of the diaphragm using an open capsule, and have a cardioid or bidirectional pickup pattern.

- There are also two main techniques for measuring the movement of the diaphragm. **Condenser microphones** use electrical capacitance between the diaphragm and a metal backplate to convert movement into electrical voltage. **Dynamic microphones** use electromagnetic induction, with a coil mounted on the back of the diaphragm and a permanent magnet inside the capsule.
- Recording of the electrical signal representing a sound can be achieved by a variety of means. The two dominant recording technologies of the twentieth century were **magnetic tape**, which magnetizes a moving strip of tape using an electromagnet, and **vinyl records**, which capture the sound waveform in the pattern of wiggles in a groove cut on the surface of a plastic disk. Both of these technologies are now obsolete, having been replaced by digital recording techniques, which are discussed in the next chapter.
- Electrical signals can be turned back into sound using a **loudspeaker**. Loudspeakers are the weak link in the sound reproduction chain: it is almost impossible to make a single loudspeaker that can faithfully project a strong sound at any frequency in any direction, and the design of speaker systems involves a number of compromises.
- By far the most common type of loudspeaker is the **dynamic loudspeaker**, which operates like a dynamic microphone in reverse: an electrical signal passed through a coil produces a magnetic field that is attracted and repelled by a nearby permanent magnet, causing the coil to vibrate. The coil is attached to a diaphragm or cone which thus vibrates as well, producing sound.
- Depending on their size and design, different loudspeakers work well in different **frequency bands**, but no speaker works well at every frequency, so complete speakers systems normally combine two or more speakers to cover all the bases.
- There are also additional issues that must be addressed for both high- and low-frequency speakers. At **high frequencies** the sound produced by a speaker is highly directional, projecting mainly to the front of the speaker and not to the sides. This issue can be addressed by making the speaker smaller. Directionality is less of a problem at **low frequencies**, but low-frequency speakers have their own issues, suffering from lower output power because the cone moves with lower velocity and hence produces less sound pressure. This issue can be addressed either by making the speaker larger, since output power is proportional to the area of the cone, or by increasing the range of movement of the cone—also called the **excursion**—which requires careful engineering to avoid distortion.

- **Complete speaker systems** normally combine separate high-frequency speakers (called tweeters), which are small, and low-frequency speakers (called woofers) which may be either large or small. Historically, woofers were large, to give them higher output power, but recent years have seen the introduction of smaller high-excursion ones, leading to today's portable speaker systems.
- Many modern speaker systems also combine speakers in different locations to create **stereophonic** or **surround sound**—the illusion of sound coming at the listener from different directions.

Exercises

7.1 In this circuit $V_{\text{in}} = 5$ volts, $R_1 = 600\Omega$, and $R_2 = 400\Omega$:



a) What is V_{out}?
b) If we want $V_{\text{out}} = 1$ volt with R_1 still being 600Ω, what should R_2 be?

7.2 We send a pure sine-wave signal into a 6 dB-per-octave filter like the one in Fig. 7.8.

a) If the frequency of the sine wave is exactly equal to the cutoff frequency f_c of the filter, by how many decibels is the intensity of the corresponding sound decreased?
b) What about if the frequency is $2f_c$?
c) If the filter is a 12 dB-per-octave one, as in Fig. 7.12, how much is the intensity cut by at f_c?

7.3 Advanced Consider a band-pass filter like the one in Fig. 7.15 on page 229, built out of a low-pass filter and a high-pass filter one after the other, and let us denote the cutoff frequencies of the two filters by

$$f_L = \frac{1}{2\pi R_1 C_1}, \qquad f_H = \frac{1}{2\pi R_2 C_2}.$$

Show that the center frequency f_C of the band-pass filter, which is the frequency at which the filter passes the largest fraction of the input signal—the center of the peak in Fig. 7.16—is $f_C = \sqrt{f_L f_H}$.

7.4 Microphones come in two main types, pressure microphones and pressure-gradient microphones.

a) How do the two types differ in terms of the direction in which they pick up sound?

If you have a choice between these two microphone types, which would be better in each of the following applications?

b) The microphone on a laptop used for videoconference calls.
c) A microphone used on a guitar amplifier for live performance.
d) A microphone used for recording vocals in the studio.
e) A headset microphone for an operator in a call center.
f) A microphone used for recording a full symphony orchestra in a concert hall.

7.5 A cardioid microphone picks up sound most strongly from the direction directly in front of the diaphragm—the 0° point in a polar plot like Fig. 7.25 on page 241. By now many decibels will the sound level drop off at (a) 45°, (b) 90°, and (c) 135°?

7.6 A hypercardioid microphone has an open-backed capsule like a cardioid microphone, but the diaphragm is not at the front of the capsule. Instead it is $\frac{1}{3}$ of the way back.

a) Show that the net sound pressure felt by a hypercardioid mic for sound coming from an angle θ off-axis is proportional to $\cos\theta + \frac{1}{3}$.
b) Hence show that the microphone picks up no sound at all from an angle of approximately 109.5° off-axis.

7.7 Advanced Either by hand or using a plotting program, make polar plots of the directional response of a cardioid microphone in the high-frequency region where Eq. (7.54) applies and (a) $fd/c = \frac{1}{2}$, (b) $fd/c = 1$.

7.8 Consider a condenser microphone of the kind discussed in Section 7.3.5.

a) What is the normal voltage of the electricity that such a microphone runs on?
b) How large is the typical separation between the diaphragm and backplate of such a microphone?
c) The typical voltage output by a condenser microphone (before amplification) is about a tenth of a volt. Hence, about how far does the diaphragm move back and forth as sound hits it?

7.9 For a condenser microphone it is important, as discussed in Section 7.3.5, that the capacitor formed by the diaphragm and backplate charge up only slowly—more slowly than the rate at which the diaphragm vibrates. In practice, this means that the time constant RC of the resistor-microphone combination in the circuit of Fig. 7.30a on page 247 must be longer than the period of the lowest-frequency sound waves. A typical capacitance for a condenser microphone is 50 picofarads. Hence what value should one use for the resistor?

7.10 A professional analog tape recorder, of the kind widely used in recording studios in the mid 20th century, has a gap of 5 μm between the poles of the tape head. At what speed should the tape move in order to fully capture the entire frequency range of audible sound?

7.11 The earliest commercial vinyl records were 10 inches in diameter and rotated at 78 revolutions per minute. The playing area extended from the rim of the record inward about 2.5

inches, which is about 6.4 cm. If the recording is to have a dynamic range of 50 dB, approximately what would be the maximum playing time of such a record?

7.12 A portable radio contains only one loudspeaker, which is 10 cm across. A listener listens to music on the radio from a point not directly in front of the speaker but 45° off axis.

a) Up to about what frequency will they hear the sound clearly?
b) Will this matter from a musical point of view?

7.13 As we have seen, loudspeakers come in a range of sizes.

a) Explain briefly the advantages and disadvantages of large loudspeakers versus small ones, paying particular attention to the relative merits of each for reproducing high and low frequencies.

With these issues in mind, approximately what size speakers should one use for the following?

b) A "tweeter" intended for frequencies above 5000 Hz.
c) A speaker for a bass amplifier that needs to capture frequencies up to 500 Hz.
d) The "driver" speaker in a set of over-the-ear headphones.

7.14 A line array speaker, like the one pictured on page 272, projects most of its sound into a narrow horizontal sheet, with only a small vertical angle and little sound going up or down. If a particular line array is 2.5 meters high, about how wide will that angle be in degrees, at 1000 Hz?