The Physics of Sound

Sound lies at the very center of speech communication. A sound wave is both the end product of the speech production mechanism and the primary source of raw material used by the listener to recover the speaker's message. Because of the central role played by sound in speech communication, it is important to have a good understanding of how sound is produced, modified, and measured. The purpose of this chapter will be to review some basic principles underlying the physics of sound, with a particular focus on two ideas that play an especially important role in both speech and hearing: the concept of the **spectrum** and **acoustic filtering**. The speech production mechanism is a kind of assembly line that operates by generating some relatively simple sounds consisting of various combinations of buzzes, hisses, and pops, and then filtering those sounds by making a number of fine adjustments to the tongue, lips, jaw, soft palate, and other articulators. We will also see that a crucial step at the receiving end occurs when the ear breaks this complex sound into its individual frequency components in much the same way that a prism breaks white light into components of different optical frequencies. Before getting into these ideas it is first necessary to cover the basic principles of vibration and sound propagation.

Sound and Vibration

A sound wave is an air pressure disturbance that results from vibration. The vibration can come from a tuning fork, a guitar string, the column of air in an organ pipe, the head (or rim) of a snare drum, steam escaping from a radiator, the reed on a clarinet, the diaphragm of a loudspeaker, the vocal cords, or virtually anything that vibrates in a frequency range that is audible to a listener (roughly 20 to 20,000 cycles per second for humans). The two conditions that are required for the generation of a sound wave are a vibratory disturbance and an elastic medium, the most familiar of which is air. We will begin by describing the characteristics of vibrating objects, and then see what happens when vibratory motion occurs in an elastic medium such as air. We can begin by examining a simple vibrating object such as the one shown in Figure 3-1. If we set this object into vibration by tapping it from the bottom, the bar will begin an upward and downward oscillation until the internal resistance of the bar causes the vibration to cease.

The graph to the right of Figure 3-1 is a visual representation of the upward and downward motion of the bar. To see how this graph is created, imagine that we use a strobe light to take a series of snapshots of the bar as it vibrates up and down. For each snapshot, we measure the **instantaneous displacement** of the bar, which is the difference between the position of the bar at the split second that the snapshot is taken and the position of the bar at rest. The rest position of the bar is arbitrarily given a displacement of zero; positive numbers are used for displacements above the rest position, and negative numbers are used for displacements below the rest position. So, the first snapshot, taken just as the bar is struck, will show an instantaneous displacement of zero; the next snapshot will show a small positive displacement, the next will show a somewhat larger positive displacement, and so on. The pattern that is traced out has a very specific shape to it. The type of vibratory motion that is produced by a simple vibratory system of this kind is called **simple harmonic motion** or **uniform circular motion**, and the pattern that is traced out in the graph is called a **sine wave** or a **sinusoid**.

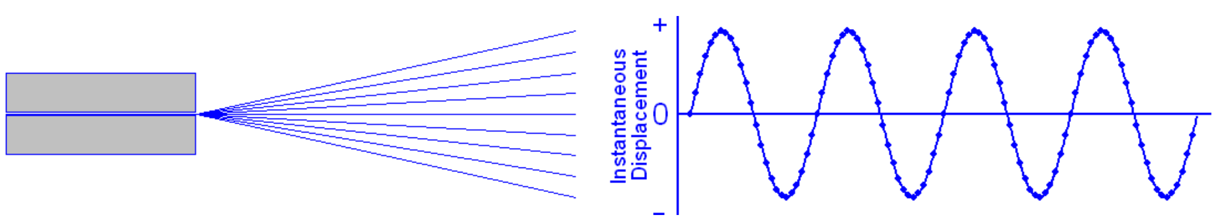


Figure 3-1. A bar is fixed at one and is set into vibration by tapping it from the bottom. Imagine that a strobe light is used to take a series of snapshots of the bar as it vibrates up and down. At each snapshot the **instantaneous displacement** of the bar is measured. Instantaneous displacement is the distance between the rest position of the bar (defined as zero displacement) and its position at any particular instant in time. Positive numbers signify displacements that are above the rest position, while negative numbers signify displacements that are below the rest position. The vibratory pattern that is traced out when the sequence of displacements is graphed is called a **sinusoid**.

Basic Terminology

We are now in a position to define some of the basic terminology that applies to sinusoidal vibration.

periodic: The vibratory pattern in Figure 3-1, and the waveform that is shown in the graph, are examples of **periodic** vibration, which simply means that there is a pattern that repeats itself over time.

cycle: **Cycle** refers to one repetition of the pattern. The instantaneous displacement waveform in Figure 3-1 shows four cycles, or four repetitions of the pattern.

period: **Period** is the time required to complete one cycle of vibration. For example, if 20 cycles are completed in 1 second, the period is 1/20th of a second (s), or 0.05 s. For speech applications, the most commonly used unit of measurement for period is the millisecond (ms):

1 ms = 1/1,000 s = 0.001 s = 10^{-3} s

A somewhat less commonly used unit is the microsecond (μs):

1 μs = 1/1,000,000 s = 0.000001 s = 10^{-6} s

frequency: Frequency is defined as the number of cycles completed in one second. The unit of measurement for frequency is **hertz** (**Hz**), and it is fully synonymous the older and more straightforward term **cycles per second** (**cps**). Conceptually, frequency is simply the rate of vibration. The most crucial function of the auditory system is to serve as a frequency analyzer – a system that determines how much energy is present at different signal frequencies. Consequently, frequency is the single most important concept in hearing science. The formula for frequency is:

f = 1/t, where: f = frequency in Hz
t = period ***in seconds***

So, for a period 0.05 s:

f = 1/t = 1/0.05 = 20 Hz

It is important to note that period must be represented in seconds in order to get the answer to come out in cycles per second, or Hz. If the period is represented in milliseconds, which is very often the case, the period first has to be converted from milliseconds into seconds by shifting the decimal point three places to the left. For example, for a period of 10 ms:

f = 1/10 ms = 1/0.01 s = 100 Hz

Similarly, for a period of 100 μs:

f = 1/100 μs = 1/0.0001 s = 10,000 Hz

The period can also be calculated if the frequency is known. Since period and frequency are inversely related, t = 1/f. So, for a 200 Hz frequency, t = 1/200 = 0.005 s = 5 ms.

Characteristics of Simple Vibratory Systems

Simple vibratory systems of this kind can differ from one another in just three dimensions: frequency, amplitude, and phase. Figure 3-2 shows examples of signals that differ in frequency. The term **amplitude** is a bit different from the other terms that have been discussed thus far, such as force and pressure. As we saw in the last chapter, terms such as force and pressure have quite specific definitions as various combinations of the basic dimensions of mass, time, and distance. Amplitude, on the other hand, will be used in this text as a generic term meaning "how much." How much what? The term amplitude can be used to refer to the magnitude of displacement, the magnitude of an air pressure disturbance, the magnitude of a force, the magnitude of power, and so on. In the present context, the term amplitude refers to the magnitude of the displacement pattern. Figure 3-3 shows two displacement waveforms that differ in amplitude. Although the concept of amplitude is as straightforward as the two waveforms shown in the figure suggest, measuring amplitude is not as simple as it might seem. The reason is that the **instantaneous amplitude** of the waveform (in this case, the displacement of the object at a particular split second in time) is constantly changing. There are many ways to measure amplitude, but a very simple method called peak-to-peak amplitude will serve our purposes well enough. Peak-to-peak amplitude is simply the difference in amplitude between the maximum positive and maximum negative peaks in the signal. For example, the bottom panel in Figure 3-3 has a peak-to-peak amplitude of 10 cm, and the top panel has a peak-to-peak amplitude of 20 cm. Figure 3-4 shows several signals that are identical in frequency and amplitude, but differ from one another in phase. The waveform labeled 0^o phase would be produced if the bar were set into vibration by tapping it from the bottom. The waveform labeled 180^o phase would be produced if the bar were set into vibration by tapping it from the top, so that the initial movement of the bar was downward rather than upward. The waveforms labeled 90^o phase and 270^o phase would be produced if the bar were set into vibration by pulling the bar to maximum displacement and letting go -- beginning at maximum positive displacement for 90^o phase, and beginning at maximum negative displacement for 270^o phase. So, the various vibratory patterns shown in Figure 3-4 are identical except with respect to phase; that is, they begin at different points in the vibratory cycle. As can be seen in Figure 3-5, the system for representing phase in degrees treats one cycle of the waveform as a circle; that is, one cycle equals 360^o. For example, a waveform that begins at zero displacement and shows its initial movement upward has a phase of 0^o, a waveform that begins at maximum positive displacement and shows its initial movement downward has a phase of 90^o, and so on.

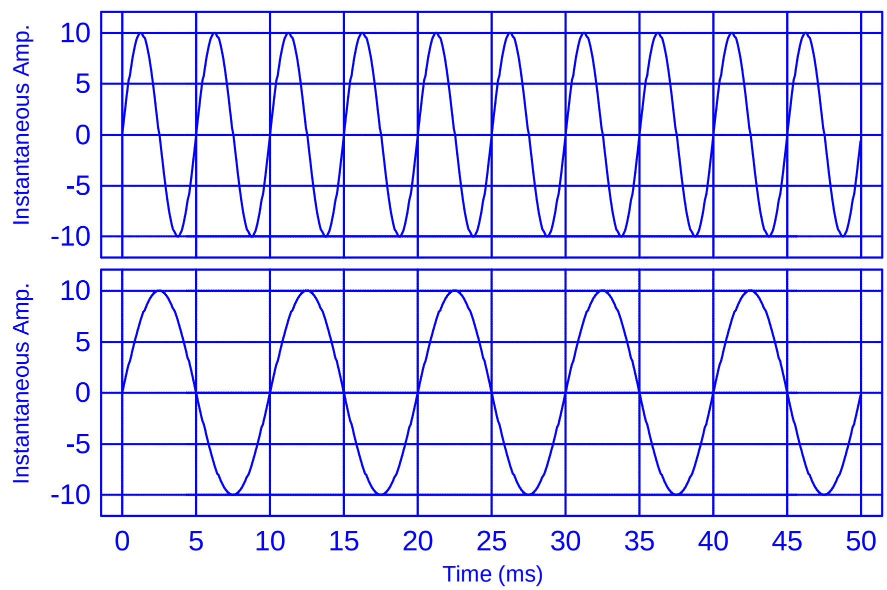


Figure 3-2. Two vibratory patterns that differ in frequency. The panel on top is higher in frequency than the panel on bottom.

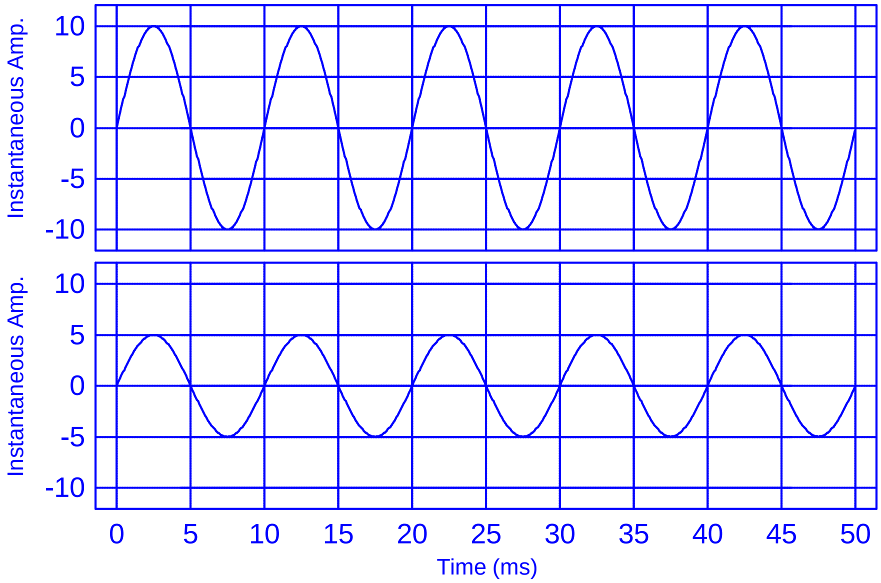


Figure 3-3. Two vibratory patterns that differ in amplitude. The panel on top is higher in amplitude than the panel on bottom.

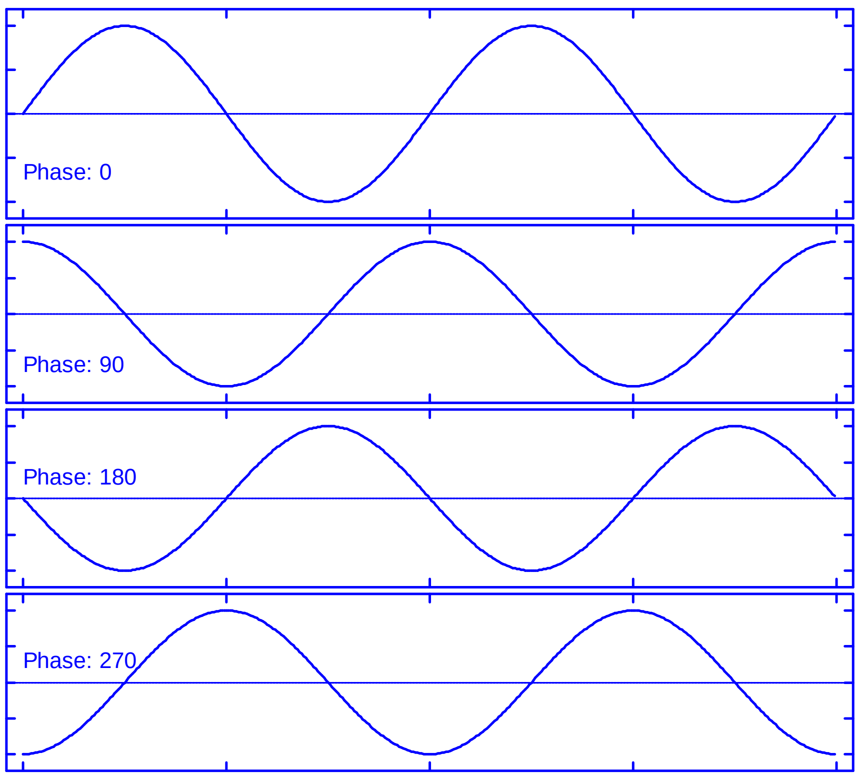


Figure 3-4. Four vibratory patterns that differ in phase. Shown above are vibratory patterns with phases of 0^0, 90^0, 180^0, and 270^0.

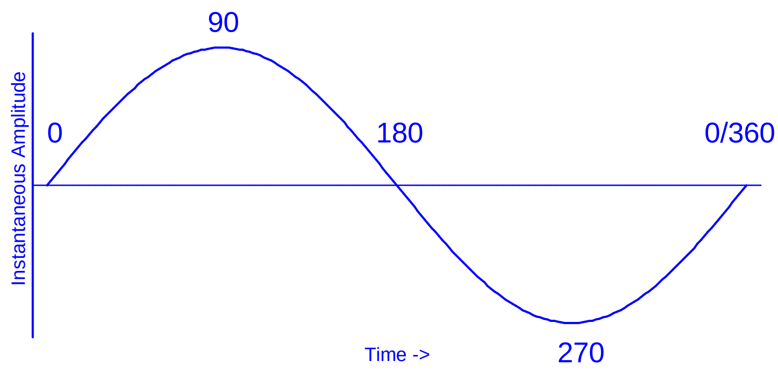


Figure 3-5. The system for representing phase treats one cycle of the vibratory pattern as a circle, consisting of 360^0. A pattern that begins at zero amplitude heading toward positive values (i.e., heading upward) is designated 0^0 phase; a waveform that begins at maximum positive displacement and shows its initial movement downward has a phase of 90^o; a waveform that begins at zero and heads downward has a phase of 180^o; and a waveform that begins at maximum negative displacement and shows its initial movement upward has a phase of 270^o. . The four phase angles that are shown above are just examples. An infinite variety of phase angles are possible.

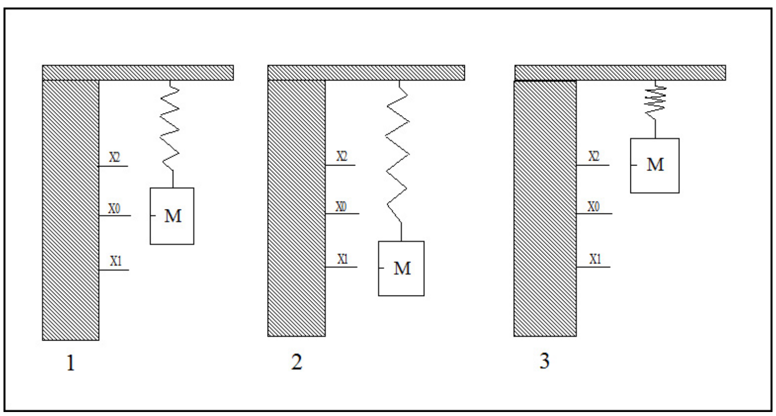


Figure 3-6. A spring and mass system whose natural vibrating frequency is controlled by two parameters: (1) the stiffness of the spring (the stiffer the spring the higher the natural vibrating frequency), and (2) the mass of the material that is suspended from the spring (the greater the mass, the lower the natural vibrating frequency).

Springs and Masses

We have noted that objects can vibrate at different frequencies, but so far have not discussed the physical characteristics that are responsible for variations in frequency. There are many factors that affect the natural vibrating frequency of an object, but among the most important are the **mass** and **stiffness** of the object. The effects of mass and stiffness on natural vibrating frequency can be illustrated with the simple spring-and-mass systems shown in Figure 3-6. In the pair of spring-and-mass systems to the left, the masses are identical but one spring is stiffer than the other. If these two spring-and-mass systems are set into vibration, the system with the stiffer spring will vibrate at a higher frequency than the system with the looser spring. This effect is similar to the changes in frequency that occur when a guitarist turns the tuning key clockwise or counterclockwise to tune a guitar string by altering its stiffness.[1]

The spring-and-mass systems to the right have identical springs but different masses. When these systems are set into vibration, the system with the greater mass will show a lower natural vibrating frequency. The reason is that the larger mass shows greater inertia and, consequently, shows greater opposition to changes in direction. Anyone who has tried to push a car out of mud or snow by rocking it back and forth knows that this is much easier with a light car than a heavy car. The reason is that the more massive car shows greater opposition to changes in direction.

In summary, the natural vibrating frequency of a spring-and-mass system is controlled by mass and stiffness. Frequency is directly proportional to stiffness (S↑F↑) and inversely proportional to mass (M↑F↓). It is important to recognize that these rules apply to all objects, and not just simple spring-and-mass systems. For example, we will see that the frequency of vibration of the vocal folds is controlled to a very large extent by muscular forces that act to alter the mass and stiffness of the folds. We will also see that the frequency analysis that is carried out by the inner ear depends to a large extent on a tuned membrane whose stiffness varies systematically from one end of the cochlea to the other.

Sound Propagation

As was mentioned at the beginning of this chapter, the generation of a sound wave requires not only vibration, but also an elastic medium in which the disturbance created by that vibration can be transmitted (see Box 3-1 [*bell jar experiment described in Patrick's science book - not yet written*]). To say that air is an elastic medium means that air, like all other matter, tends to return to its original shape after it is deformed through the application of a force. The prototypical example of an object that exhibits this kind of restoring force is a spring. To understand the mechanism underlying sound propagation, it is useful to think of air as consisting of collection of particles that are connected to one another by springs, with the springs representing the restoring forces associated with the elasticity of the medium. **Air pressure** is related to particle **density**. When a volume of air is undisturbed, the individual particles of air distribute themselves more-or-less evenly, and the elastic forces are at their resting state. A volume of air that is in this undisturbed state it is said to be at **atmospheric pressure**. For our purposes, atmospheric pressure can be defined in terms of two interrelated conditions: (1) the air molecules are approximately evenly spaced, and (2) the elastic forces, represented by the interconnecting springs, are neither compressed nor stretched beyond their resting state. When a vibratory disturbance causes the air particles to crowd together (i.e., producing an increase in particle density), air pressure is higher than atmospheric, and the elastic forces are in a **compressed** state. Conversely, when particle spacing is relatively large, air pressure is lower than atmospheric.

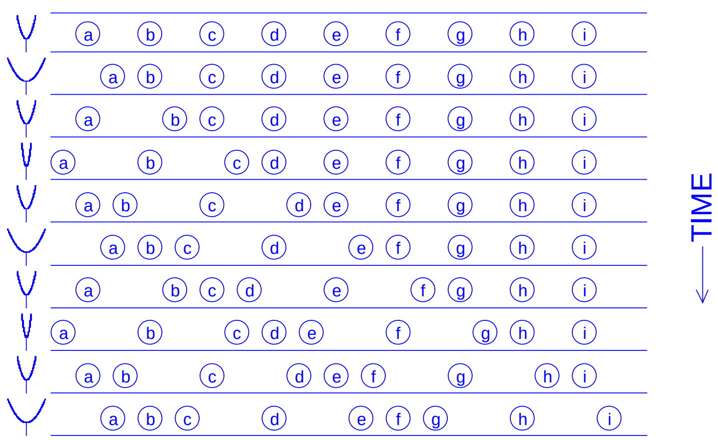


Figure 3-7. Shown above is a highly schematic illustration of the chain reaction that results in the propagation of a sound wave (modeled after Denes and Pinson, 1963).

[1]The example of tuning a guitar string is imperfect since the mass of the vibrating portion of the string decreases slightly as the string is tightened. This occurs because a portion of the string is wound onto the tuning key as it is tightened.

When a vibrating object is placed in an elastic medium, an air pressure disturbance is created through a chain reaction similar to that illustrated in Figure 3-7. As the vibrating object (a tuning fork in this case) moves to the right, particle *a*, which is immediately adjacent to the tuning fork, is displaced to the right. The elastic force generated between particles *a* and *b* (not shown in the figure) has the effect a split second later of displacing particle b to the right. This disturbance will eventually reach particles c, d, e, and so on, and in each case the particles will be momentarily crowded together. This crowding effect is called **compression** or **condensation**, and it is characterized by dense particle spacing and, consequently, air pressure that is slightly higher than atmospheric pressure. The propagation of the disturbance is analogous to the chain reaction that occurs when an arrangement of dominos is toppled over. Figure 3-7 also shows that at some close distance to the left of a point of compression, particle spacing will be greater than average, and the elastic forces will be in a stretched state. This effect is called **rarefaction**, and it is characterized by relatively wide particle spacing and, consequently, air pressure that is slightly lower than atmospheric pressure.

The compression wave, along with the rarefaction wave that immediately follows it, will be propagated outward at the speed of sound. The speed of sound varies depending on the average elasticity and density of the medium in which the sound is propagated, but a good working figure for air is about 35,000 centimeters per second, or approximately 783 miles per hour. Although Figure 3-7 gives a reasonably good idea of how sound propagation works, it is misleading in two respects. First, the scale is inaccurate to an absurd degree: a single cubic inch of air contains approximately 400 billion molecules, and not the handful of particles shown in the figure. Consequently, the compression and rarefaction effects are statistical rather than strictly deterministic as shown in Figure 3-7. Second, although Figure 3-7 makes it appear that the air pressure disturbance is propagated in a simple straight line from the vibrating object, it actually travels in all directions from the source. This idea is captured somewhat better in Figure 3-8, which shows sound propagation in two of the three dimensions in which the disturbance will be transmitted. The figure shows rod and piston connected to a wheel spinning at a constant speed. Connected to the piston is a balloon that expands and contracts as the piston moves in and out of the cylinder. As the balloon expands the air particles are compressed; i.e., air pressure is momentarily higher than atmospheric. Conversely, when the balloon contracts the air particles are sucked inward, resulting in rarefaction. The alternating compression and rarefaction waves are propagated outward in all directions form the source. Only two of the three dimensions are shown here; that is, the shape of the pressure disturbance is actually spherical rather than the circular pattern that is shown here. Superimposed on the figure, in the graph labeled “one line of propagation,” is the resulting air pressure waveform. Note that the pressure waveform takes on a high value during instants of compression and a low value during instants of rarefaction. The figure also gives some idea of where the term **uniform circular motion** comes from. If one were to make a graph plotting the height of the connecting rod on the rotating wheel as a function of time it would trace out a perfect sinusoid; i.e., with exactly the shape of the pressure waveform that is superimposed on the figure.

The Sound Pressure Waveform

Returning to Figure 3-7 for a moment, imagine that we chose some specific distance from the tuning fork to observe how the movement and density of air particles varied with time. We would see individual air particles oscillating small distances back and forth, and if we monitored particle density we would find that high particle density (high air pressure) would be followed a moment later by relatively even particle spacing (atmospheric pressure), which would be followed by a moment later by wide particle spacing (low air pressure), and so on. Therefore, for an object that is vibrating sinusoidally, a graph showing variations in **instantaneous air pressure** over time would also be sinusoidal. This is illustrated in Figure 3-9.

The vibratory patterns that have been discussed so far have all been sinusoidal. The concept of a sinusoid has not been formally defined, but for our purposes it is enough to know that a sinusoid has precisely the smooth shape that is shown in Figures such as 3-4 and 3-5. While sinusoids, also known as **pure tones**, have a very special place in acoustic theory, they are rarely encountered in nature. The sound produced by a tuning fork comes quite close to a sinusoidal shape, as do the simple tones that are used in hearing tests. Much more common in both speech and music are more complex, nonsinusoidal patterns, to be discussed below. As will be seen in later chapters, these complex vibratory patterns play a very important role in speech.

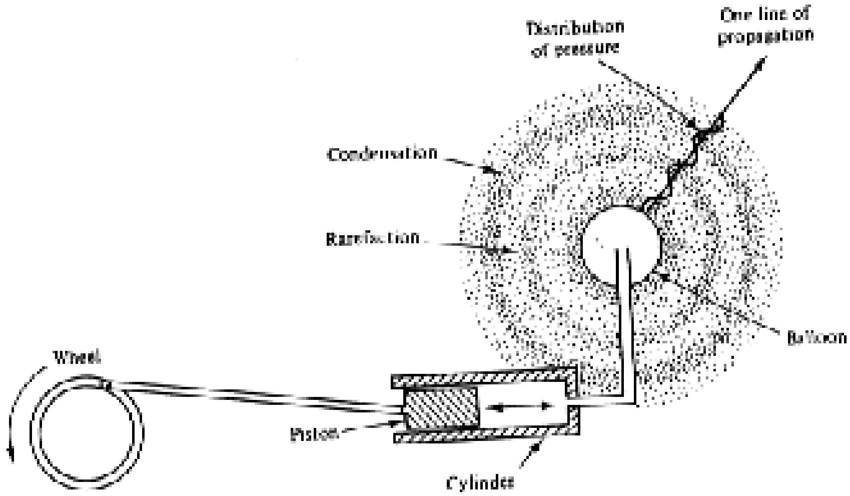


Figure 3-8 Illustration of the propagation of a sound wave in two dimensions.

The Frequency Domain

We now arrive at what is probably the single most important concept for understanding both hearing and speech acoustics. The graphs that we have used up to this point for representing either vibratory motion or the air pressure disturbance created by this motion are called **time domain representations**. These graphs show how instantaneous displacement (or instantaneous air pressure) varies over time. Another method for representing either sound or vibration is called a **frequency domain representation**, also known as a **spectrum**. There are, in fact, two kinds of frequency domain representations that are used to characterize sound. One is called an **amplitude spectrum** (also known as a **magnitude spectrum** or a **power spectrum,** depending on how the level of the signal is represented) and the other is called a **phase spectrum**. For reasons that will become clear soon, the amplitude spectrum is by far the more important of the two. An amplitude spectrum is simply a graph showing what frequencies are present with what amplitudes. Frequency is given along the x axis and some measure of amplitude is given on the y axis. A phase spectrum is a graph showing what frequencies are present with what phases.

Figure 3-10 shows examples of the amplitude and phase spectra for several sinusoidal signals. The top panel shows a time-domain representation of a sinusoid with a period of 10 ms and, consequently, a frequency of 100 Hz ($f = 1/t = 1/0.01$ sec = 100 Hz). The peak-to-peak amplitude for this signal is 400 μPa, and the signal has a phase of 90^o. Since the amplitude spectrum is a graph showing what frequencies are present with what amplitudes, the amplitude spectrum for this signal will show a single line at 100 Hz with a height of 400 μPa. The phase spectrum is a graph showing what frequencies are present with what phases, so the phase spectrum for this signal will show a single line at 100 Hz with a height of 90^o. The second panel in Figure 3-10 shows a 200 Hz sinusoid with a peak-to-peak amplitude of 200 μPa and a phase of 180^o. Consequently, the amplitude spectrum will show a single line at 200 Hz with a height of 100 μPa, while the phase spectrum will show a line at 200 Hz with a height of 180^o.

Complex Periodic Sounds

Sinusoids are sometimes referred to as **simple periodic** signals. The term "periodic" means that there is a pattern that repeats itself, and the term "simple" means that there is only one frequency component present. This is confirmed in the frequency domain representations in Figure 3-10, which all show a single frequency component in both the amplitude and phase spectra. **Complex periodic** signals involve the repetition of a nonsinusoidal pattern, and in all cases, complex periodic signals consist of more than a single frequency component. *All nonsinusoidal periodic signals are considered complex periodic.*

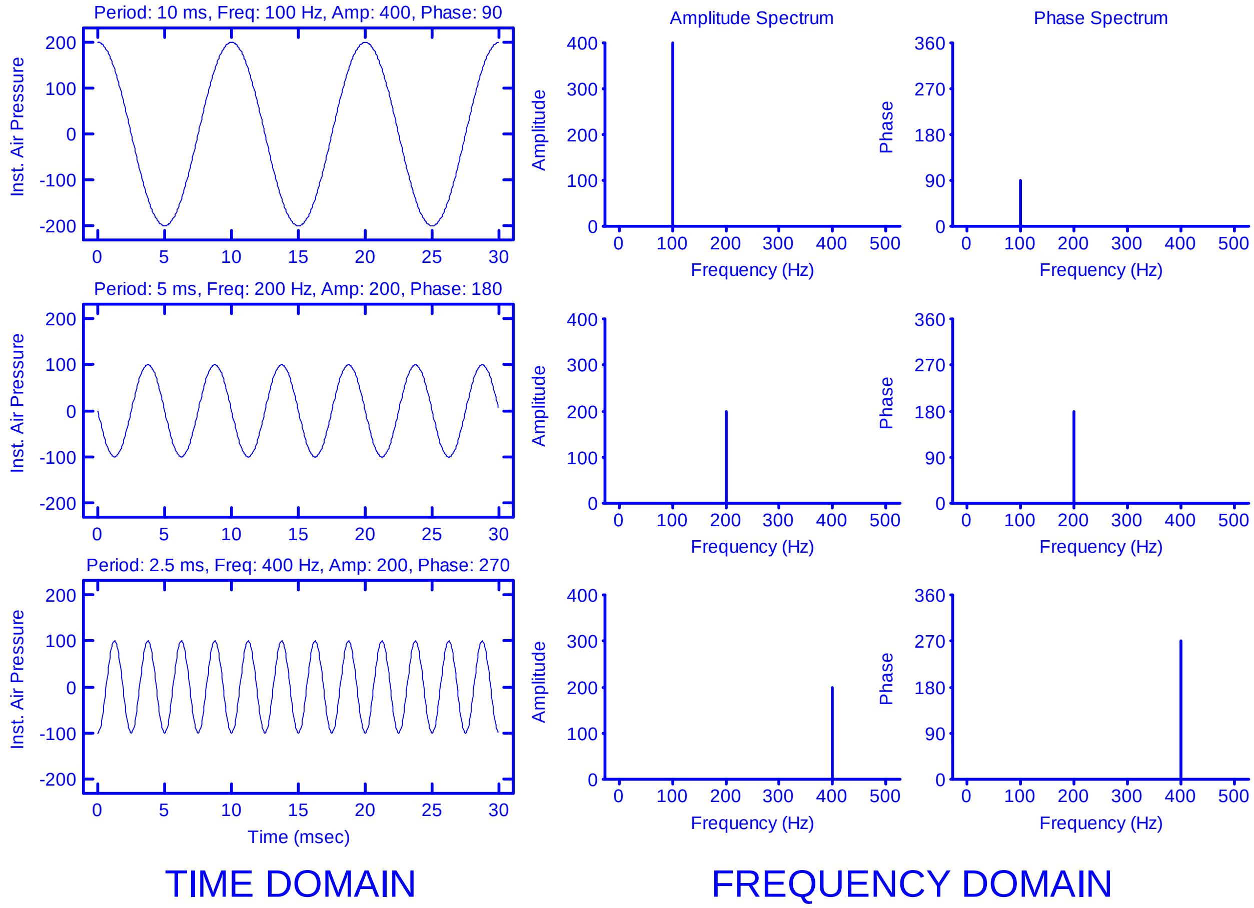


Figure 3-10. Time and frequency domain representations of three sinusoids. The frequency domain consists of two graphs: an amplitude spectrum and a phase spectrum. An amplitude spectrum is a graph showing what frequencies are present with what amplitudes, and a phase spectrum is a graph showing the phases of each frequency component.

Figure 3-11 shows several examples of complex periodic signals, along with the amplitude spectra for these signals. The time required to complete one cycle of the complex pattern is called the **fundamental period**. This is precisely the same concept as the term **period** that was introduced earlier. The only reason for using the term "fundamental period" instead of the simpler term "period" for complex periodic signals is to differentiate the fundamental period (the time required to complete one cycle of the pattern as a whole) from other periods that may be present in the signal (e.g., more rapid oscillations that might be observed within each cycle). The symbol for fundamental period is t_o. **Fundamental frequency** (f_o) is calculated from fundamental period using the same kind of formula that we used earlier for sinusoids:

$$f_o = 1/t_o$$

The signal in the top panel of Figure 3-11 has a fundamental period of 5 ms, so $f_o = 1/0.005 = 200$ Hz.

Examination of the amplitude spectra of the signals in Figure 3-11 confirms that they do, in fact, consist of more than a single frequency. In fact, complex periodic signals show a very particular kind of amplitude spectrum called a **harmonic spectrum**. A harmonic spectrum shows energy at the fundamental frequency *and at whole number multiples of the fundamental frequency*. For example, the signal in the top panel of Figure 3-11 has energy present at 200 Hz, 400 Hz, 600 Hz, 800 Hz, 1,000 Hz, 1200 Hz, and so on. Each frequency component in the

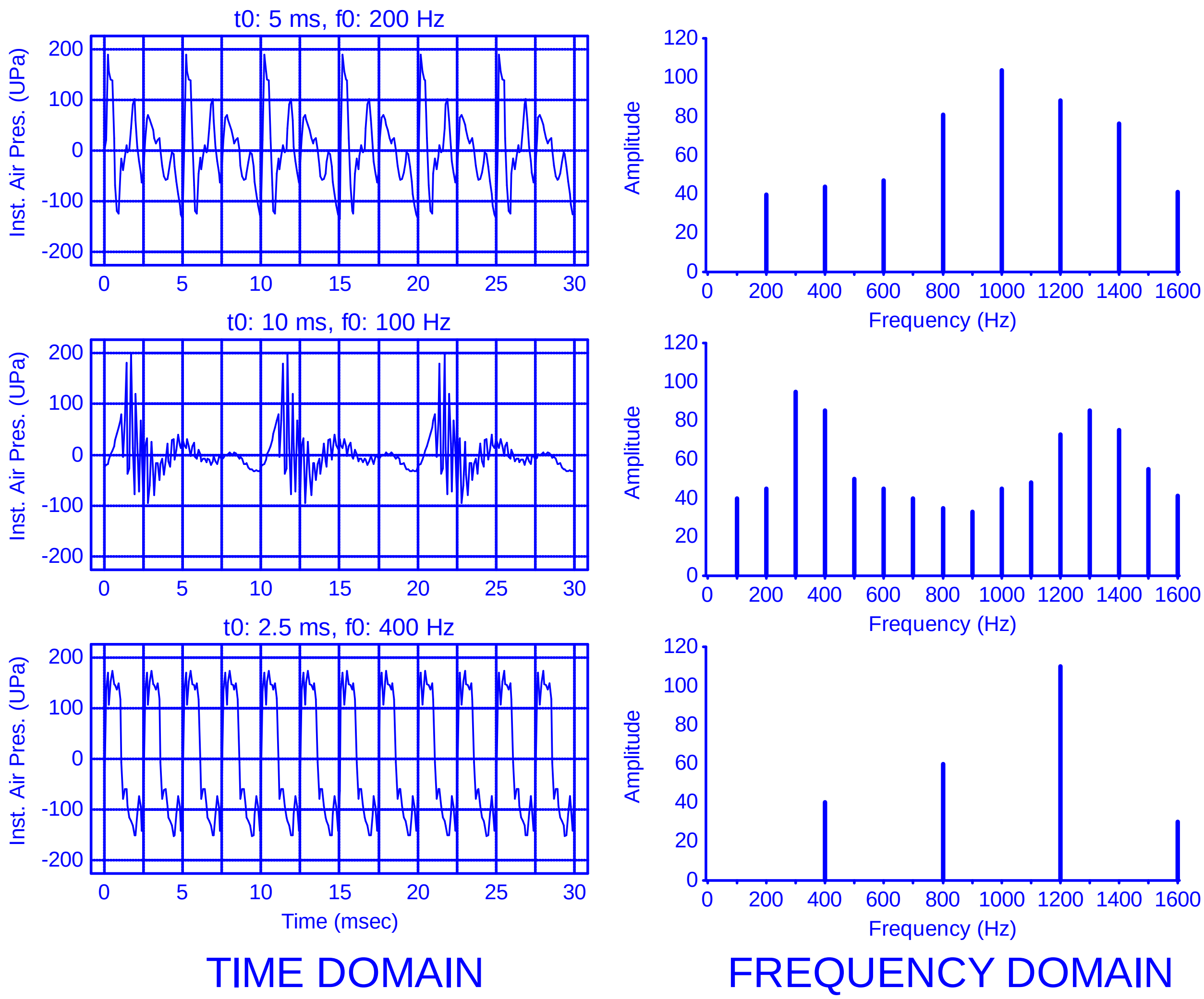


Figure 3-11. Time and frequency domain representations of three complex periodic signals. Complex periodic signals have harmonic spectra, with energy at the fundamental frequency (f_0) and at whole number multiples of f_0 ($f_0 \cdot 2$, $f_0 \cdot 3$, $f_0 \cdot 4$, etc.) For example, the signal in the upper left, with a fundamental frequency of 200 Hz, shows energy at 200 Hz, 400 Hz, 600 Hz, etc. In the spectra on the right, amplitude is measured in arbitrary units. The main point being made in this figure is the distribution of harmonic frequencies at whole number multiples of f_0 for complex periodic signals.

amplitude spectrum of a complex periodic signal is called a **harmonic** (also known as a **partial**). The fundamental frequency, in this case 200 Hz, is also called the first harmonic, the 400 Hz component ($2 \cdot f_o$) is called the second harmonic, the 600 Hz component ($3 \cdot f_o$) is called the third harmonic, and so on.

The second panel in Figure 3-11 shows a complex periodic signal with a fundamental period of 10 ms and, consequently, a fundamental frequency of 100 Hz. The harmonic spectrum that is associated with this signal will therefore show energy at 100 Hz, 200 Hz, 300 Hz, 400 Hz, 500 Hz, and so on. The bottom panel of Figure 3-11 shows a complex periodic signal with a fundamental period of 2.5 ms, a fundamental frequency of 400 Hz, and harmonics at 400, 800, 1200, 1600, and so on. Notice that there two completely interchangeable ways to define the term fundamental frequency. In the time domain, the fundamental frequency is the number of cycles of the complex pattern that are completed in one second. In the frequency domain, except in the case of certain special signals, the fundamental frequency is the lowest harmonic in the harmonic spectrum. Also, the fundamental frequency defines the harmonic spacing; that is, when the fundamental frequency is 100 Hz, harmonics will be spaced at 100 Hz intervals (i.e., 100, 200, 300 ...), when the fundamental frequency is 125 Hz, harmonics will be spaced at 125 Hz intervals (i.e., 125, 250, 375...), and when the fundamental frequency is 200 Hz, harmonics will be spaced at 200 Hz intervals (i.e., 200, 400, 600 ...). (For some special signals this will not be the case.[2]) So, when f_o is low, harmonics will be closely spaced, and when f_o is high, harmonics will be widely spaced. This is clearly seen in Figure 3-11: the signal with the lowest f_0 (100 Hz, the middle signal) shows the narrowest harmonic spacing, while the signal with the highest f_0 (400 Hz, the bottom signal) shows the widest harmonic spacing.

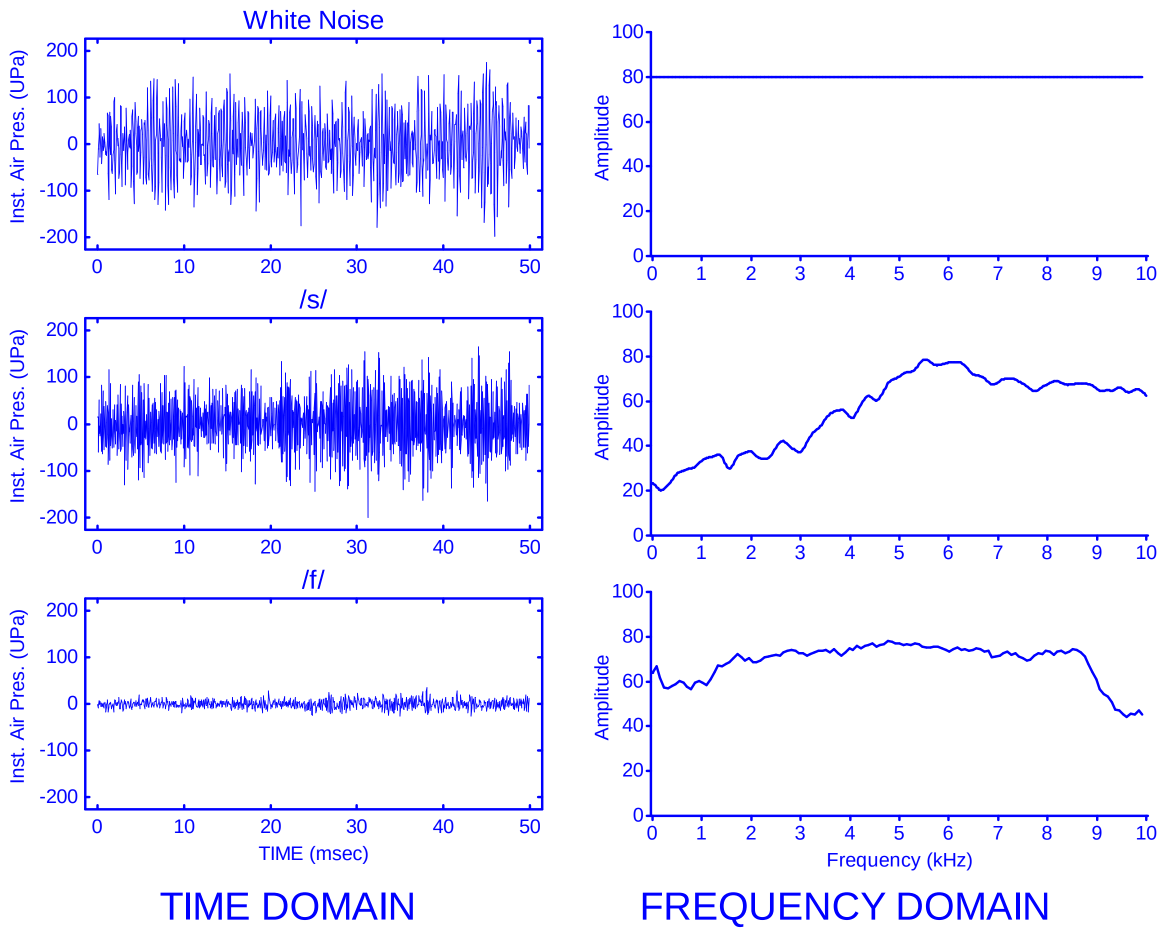


Figure 3-12. Time and frequency domain representations of three non-transient complex aperiodic signals. Unlike complex periodic signals, complex aperiodic signals show energy that is spread across the spectrum. This type of spectrum is called **dense** or **continuous**. These spectra have a very different appearance from the "picket fence" look that is associated with the **discrete**, harmonic spectra of complex periodic signals.

There are certain characteristics of the spectra of complex periodic sounds that can be determined by making simple measurements of the time domain signal, and there are certain other characteristics that require a more complex analysis. For example, simply by examining the signal in the bottom panel of Figure 3-11 we can determine that it is complex periodic (i.e., it is periodic but not sinusoidal) and therefore it will show a harmonic spectrum with energy at whole number multiples of the fundamental frequency. Further, by measuring the fundamental period (2.5 ms)

[2]There are some complex periodic signals that have energy at *odd* multiples of the fundamental frequency only. A ***square wave***, for example, is a signal that alternates between maximum positive amplitude and maximum negative amplitude. The spectrum of square wave shows energy at odd multiples of the fundamental frequency only. Also, a variety of simple signal processing tricks can be used to create signals with harmonics at any arbitrary set of frequencies. For example, it is a simple matter to create a signal with energy at 400, 500, and 600 Hz only. While these kinds of signals can be quite useful for conducting auditory perception experiments, it remains true that most naturally occurring complex periodic signals have energy at all whole number multiples of the fundamental frequency.

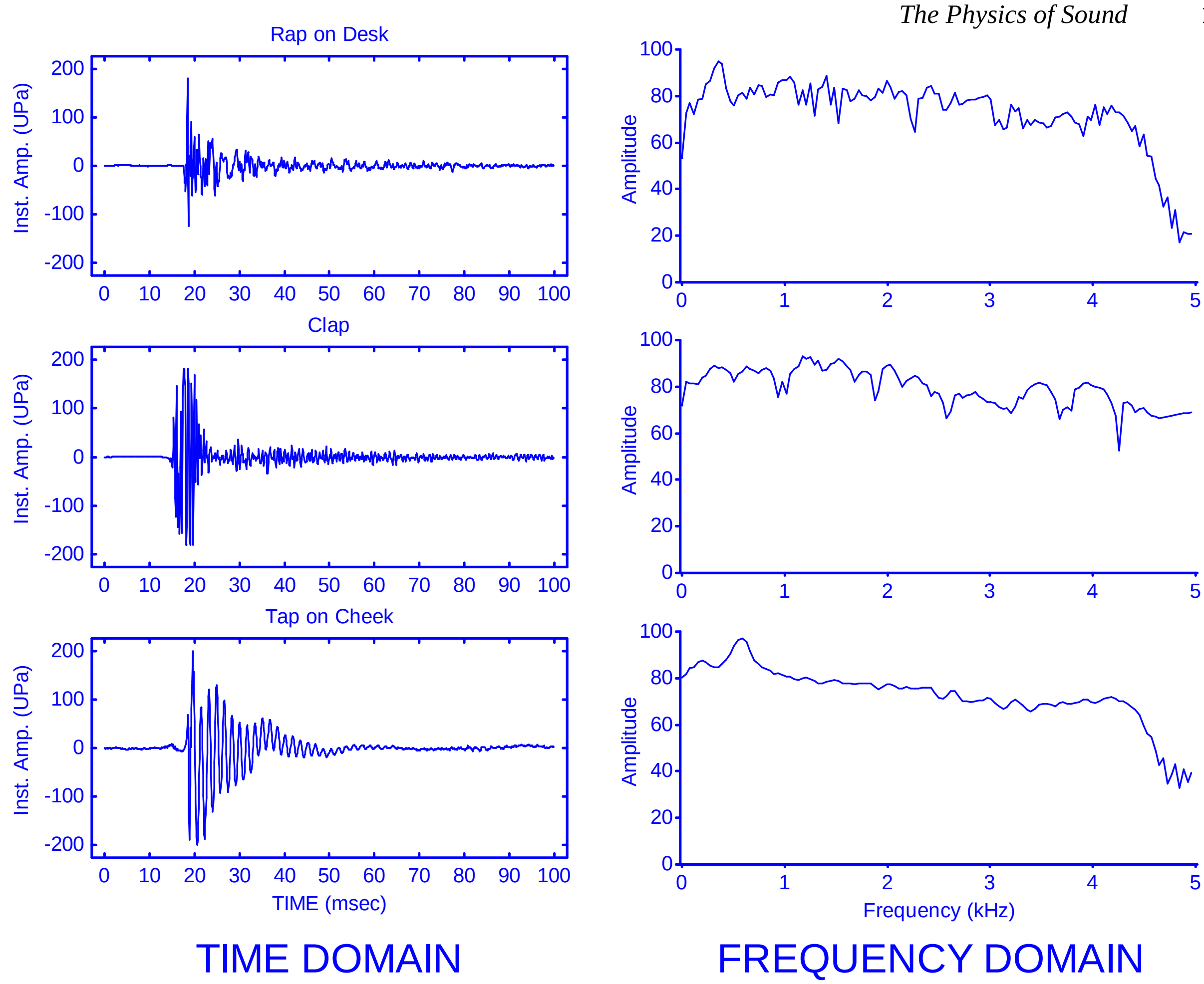


Figure 3-13. Time and frequency domain representations of three transients. Transients are complex aperiodic signals that are defined by their brief duration. Pops, clicks, and the sound gun fire are examples of transients. In common with longer duration complex aperiodic signals, transients show **dense** or **continuous** spectra, very unlike the discrete, harmonic spectra associated with complex periodic

and converting it into fundamental frequency (400 Hz), we are able to determine that the signal will have energy at 400, 800, 1200, 1600, etc. But how do we know the amplitude of each of these frequency components? And how do we know the phase of each component? The answer is that you cannot determine harmonic amplitudes or phases simply by inspecting the signal or by making simple measurements of the time domain signals with a ruler. We will see soon that a technique called **Fourier analysis** is able to determine both the amplitude spectrum and the phase spectrum of any signal. We will also see that the inner ears of humans and many other animals have developed a trick that is able to produce a neural representation that is comparable in some respects to an amplitude spectrum. We will also see that the ear has *no* comparable trick for deriving a representation that is equivalent to a phase spectrum. This explains why the amplitude spectrum is far more important for speech and hearing applications than the phase spectrum. We will return to this point later.

To summarize: (1) a complex periodic signal is any periodic signal that is not sinusoidal, (2) complex periodic signals have energy at the fundamental frequency (f_o) and at whole number multiples of the fundamental frequency ($2 \cdot f_o$, $3 \cdot f_o$, $4 \cdot f_o$ ...), and (3) although measuring the fundamental frequency allows us to determine the *frequency locations* of harmonics, there is no simple measurement that can tell us harmonic amplitudes or phases. For this, Fourier analysis or some other spectrum analysis technique is needed.

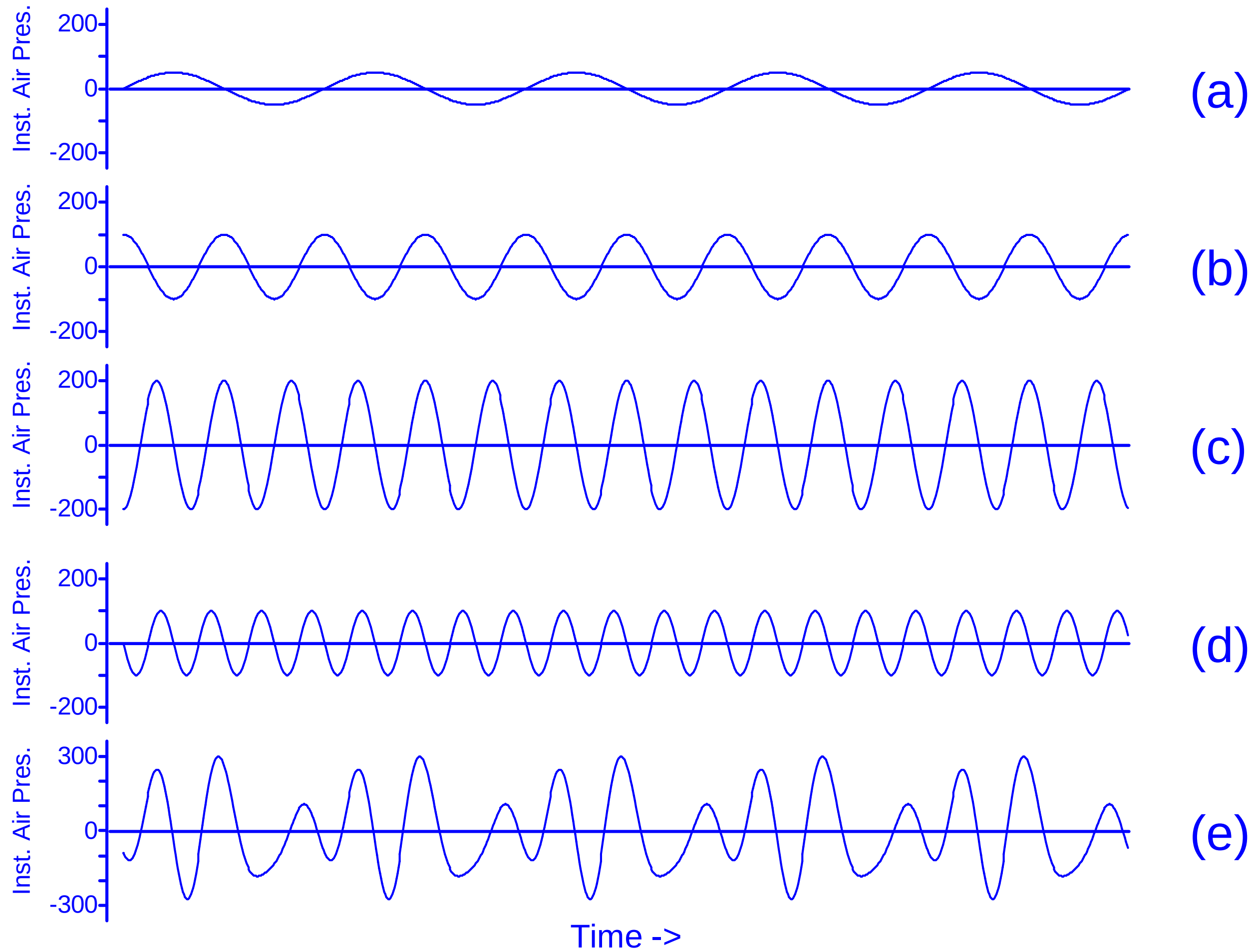


Figure 3-14. Illustration of the principle underlying Fourier analysis. The complex periodic signal shown in panel *e* was derived by point-for-point summation of the sinusoidal signals shown in panels *a-d*. Point-for-point summation simply means beginning at time zero (i.e., the start of the signal) and adding the instantaneous amplitude of signal *a* to the instantaneous amplitude of signal *b* at time zero, then adding that sum to the instantaneous amplitude of signal *c*, also at time zero, then adding that sum to instantaneous amplitude of signal *d* at time zero. The sum of instantaneous amplitudes at time zero of signals a-d is the instantaneous amplitude of the composite signal *e* at time zero. For example, at time zero the amplitudes of sinusoids a-d are 0, +100, -200, and 0, respectively, producing a sum of -100. This agrees with the instantaneous amplitude at the very beginning of composite signal e. The same summation procedure is followed for all time points.

Aperiodic Sounds

An **aperiodic** sound is any sound that does not show a repeating pattern in its time domain representation. There are many aperiodic sounds in speech. Examples include the hissy sounds associated with fricatives such as /f/ and /s/, and the various hisses and pops associated with articulatory release for the stop consonants /b,d,g,p,t,k/. Examples of non-speech aperiodic sounds include a drummer's cymbal or snare drum, the hiss produced by a radiator, and static sound produced by a poorly tuned radio. There are two types of aperiodic sounds: (1) **continuous aperiodic** sounds (also known as **noise**) and (2) **transients**. Although there is no sharp cutoff, the distinction between continuous aperiodic sounds and transients is based on duration. Transients (also "pops" and "clicks") are defined by their very brief duration, and continuous aperiodic sounds are of longer duration. Figure 3-12 shows several examples of time domain representations and amplitude spectra for continuous aperiodic sounds. The lack of periodicity in the time

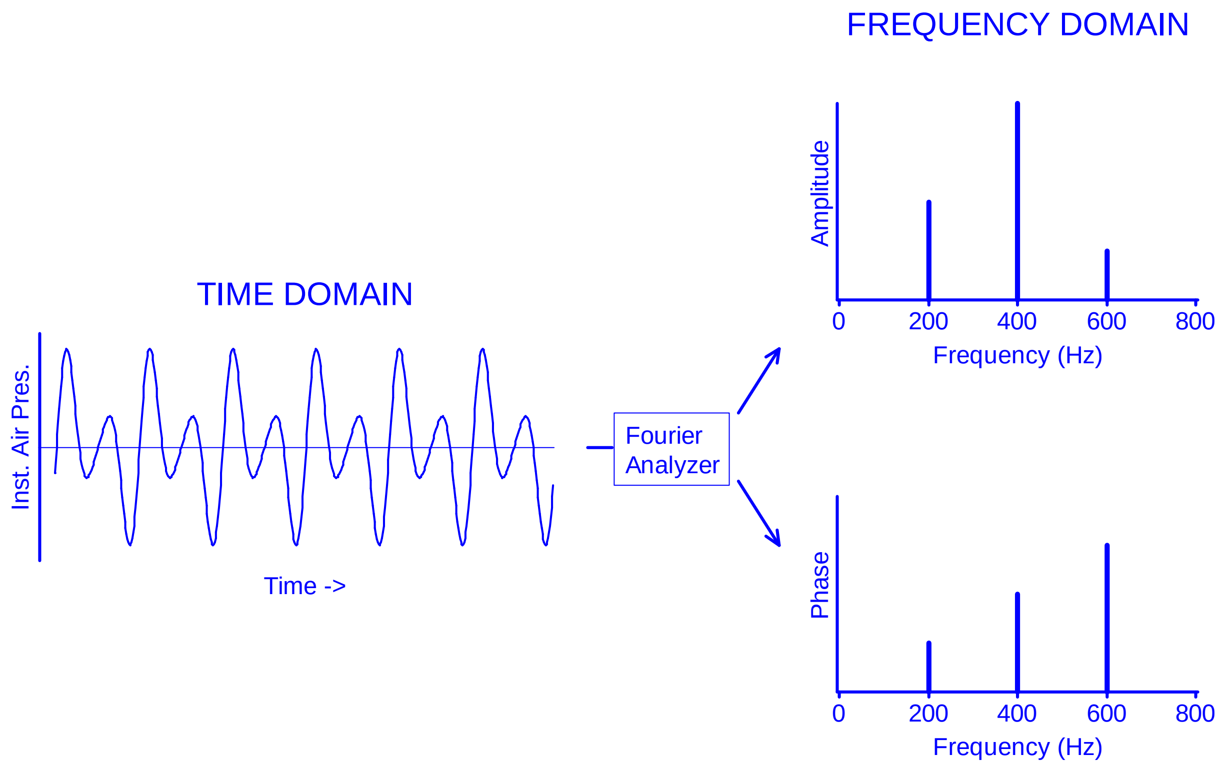


Figure 3-15. A signal enters a Fourier analyzer in the time domain and exits in the frequency domain. As outputs, the Fourier analyzer produces two frequency-domain representations: an amplitude spectrum that shows the amplitude of each sinusoidal component that is present in the input signal, and a phase spectrum that shows the phase of each of the sinusoids. The input signal can be reconstructed perfectly by summing sinusoids at frequencies, amplitudes, and phase that are shown in the Fourier amplitude and phase spectra, using the summing method that is illustrated in Figure 3-14..

domain is quite evident; that is, unlike the periodic sounds we have seen, there is no pattern that repeats itself over time.

All aperiodic sounds -- both continuous and transient -- are complex in the sense that they always consist of energy at more than one frequency. The characteristic feature of aperiodic sounds in the frequency domain is a **dense** or **continuous spectrum**, which stands in contrast to the harmonic spectrum that is associated with complex periodic sounds. In a harmonic spectrum, there is energy at the fundamental frequency, followed by a gap with little or no energy, followed by energy at the second harmonic, followed by another gap, and so on. The spectra of aperiodic sounds do not share this "picket fence" appearance. Instead, energy is smeared more-or-less continuously across the spectrum. The top panel in Figure 3-12 shows a specific type of continuous aperiodic sound called **white noise.** By analogy to white light, white noise has a flat amplitude spectrum; that is, approximately equal amplitude at all frequencies. The middle panel in Figure 3-12 shows the sound /s/, and the bottom panel shows sound /f/. Notice that the spectra for all three sounds are **dense**; that is, they do not show the "picket fence" look that reveals harmonic structure. As was the case for complex periodic sounds, there is no way to tell how much energy there will be at different frequencies by inspecting the time domain signal or by making any simple measures with a ruler. Likewise, there is no simple way to determine the phase spectrum. So, after inspecting a time-domain signal and determining that it is aperiodic, all we know for sure is that it will have a dense spectrum rather than a harmonic spectrum.

Figure 3-13 shows time domain representations and amplitude spectra for three transients. The transient in the top panel was produced by rapping on a wooden desk, the second is a single clap of the hands, and the third was produced by holding the mouth in position for the vowel /o/, and tapping the cheek with an index finger. Note the brief durations of the signals. Also, as with continuous aperiodic sounds, the spectra associated with transients are dense; that is, there is no evidence of harmonic organization. In speech, transients occur at the instant of articulatory release for stop consonants. There are also some languages, such as the South African languages Zulu, Hottentot, and Xhosa, that contain mouth clicks as part of their phonemic inventory (MacKay, 1986). **Fourier Analysis**

Fourier analysis is an extremely powerful tool that has widespread applications in nearly every major branch of physics and engineering. The method was developed by the 19^{th} century mathematician Joseph Fourier, and although Fourier was studying thermal waves at the time, the technique can be applied to the frequency analysis of any kind of wave. Fourier's great insight was the discovery that *all complex waves can be derived by adding sinusoids together*, so long as the sinusoids are of the appropriate frequencies, amplitudes, and phases. For example, the complex periodic signal at the bottom of Figure 3-14 can be derived by summing sinusoids at 100, 200, 300, and 400 Hz, with each sinusoidal component having the amplitude and phase that is shown in the figure (see the caption of Figure 3-14 for an explanation of what is meant by summing the sinusoidal components). The assumption that all complex waves can be derived by adding sinusoids together is called **Fourier's theorem**, and the analysis technique that Fourier developed from this theorem is called Fourier analysis. Fourier analysis is a mathematical technique that takes a time domain signal as its input and determines: (1) the amplitude of each sinusoidal component that is present in the input signal, and (2) the phase of each sinusoidal component that is present in the input signal. Another way of stating this is that Fourier analysis takes a time domain signal as its input and produces *two* frequency domain representations as output: (1) an amplitude spectrum, and (2) a phase spectrum.

The basic concept is illustrated in Figure 3-15, which shows a time domain signal entering the Fourier analyzer. Emerging at the output of the Fourier analyzer is an amplitude spectrum (a graph showing the amplitude of each sinusoid that is present in the input signal) and a phase spectrum (a graph showing the phase of each sinusoid that is present in the input signal). The amplitude spectrum tells us that the input signal contains: (1) 200 Hz sinusoid with an amplitude of 100 μPa, a 400 Hz sinusoid with an amplitude of 200 μPa, and a 600 Hz sinusoid with an amplitude of 50 μPa. Similarly, the phase spectrum tells us that the 200 Hz sinusoid has a phase of 90^o, the 400 Hz sinusoid has a phase of 180^o, and the 600 Hz sinusoid has a phase of 270^o. If Fourier's theorem is correct, we should be able to reconstruct the input signal by summing sinusoids at 200, 400, and 600 Hz, using the amplitudes and phases that are shown. In fact, summing these three sinusoids in this way would *precisely* reproduce the original time domain signal; that is, we would get back an exact replica of our original signal, and not just a rough approximation to it.

For our purposes it is not important to understand how Fourier analysis works. The most important point about Fourier's idea is that, visual appearances aside, all complex waves consist of sinusoids of varying frequencies, amplitudes, and phases. In fact, Fourier analysis applies not only to periodic signals such as those shown in Figure 3-15, but also to noise and transients. In fact, the amplitude spectra of the aperiodic signals shown in Figure 3-13 were calculated using Fourier analysis. In later chapters we will see that the auditory system is able to derive a neural representation that is roughly comparable to a Fourier amplitude spectrum. However, as was mentioned earlier, the auditory system does not derive a representation comparable to a Fourier phase spectrum. As a result, listeners are very sensitive to changes in the amplitude spectrum but are relatively insensitive to changes in phase.

Some Additional Terminology

Overtones vs. Harmonics: The term **overtone** and the term harmonic refer to the same concept; they are just counted differently. As we have seen, in a harmonic series such as 100, 200, 300, 400, etc., the 100 Hz component can be referred to as either the fundamental frequency or the first harmonic; the 200 Hz component is the second harmonic, the 300 Hz component is the third harmonic, and so on. An alternative set of terminology would refer to the 100 Hz component as the fundamental frequency, the 200 Hz component as the **first overtone**, the 300 Hz component as the **second overtone**, and so on. Use of the term overtone tends to be favored by those interested in musical acoustics, while most other acousticians tend to use the term harmonic.

Octaves vs. Harmonics: An **octave** refers to a doubling of frequency. So, if we begin at 100 Hz, the next octave up would 200 Hz, the next would be 400 Hz, the next would be 800 Hz, and so on. Note that this is quite different from a harmonic progression. A harmonic progression beginning at 300 Hz would be 300, 600, 900, 1200, 1500, etc., while an octave progression would be 300, 600, 1200, 2400, 4800, etc. There is something auditorilly natural about octave spacing, and octaves play a very important role in the organization of musical scales. For example, on a piano keyboard, middle A (A_5) is 440 Hz, A above middle A (A_6) is 880 Hz, A_7 is 1,760 and so on. (See Box 3-2).

Wavelength: The concept of **wavelength** is best illustrated with an example given by Small (1973). Small asks us to imagine dipping a finger repeatedly into a puddle of water at a perfectly regular interval. Each time the finger hits the water, a wave is propagated outward, and we would see a pattern formed consisting of a series of concentric

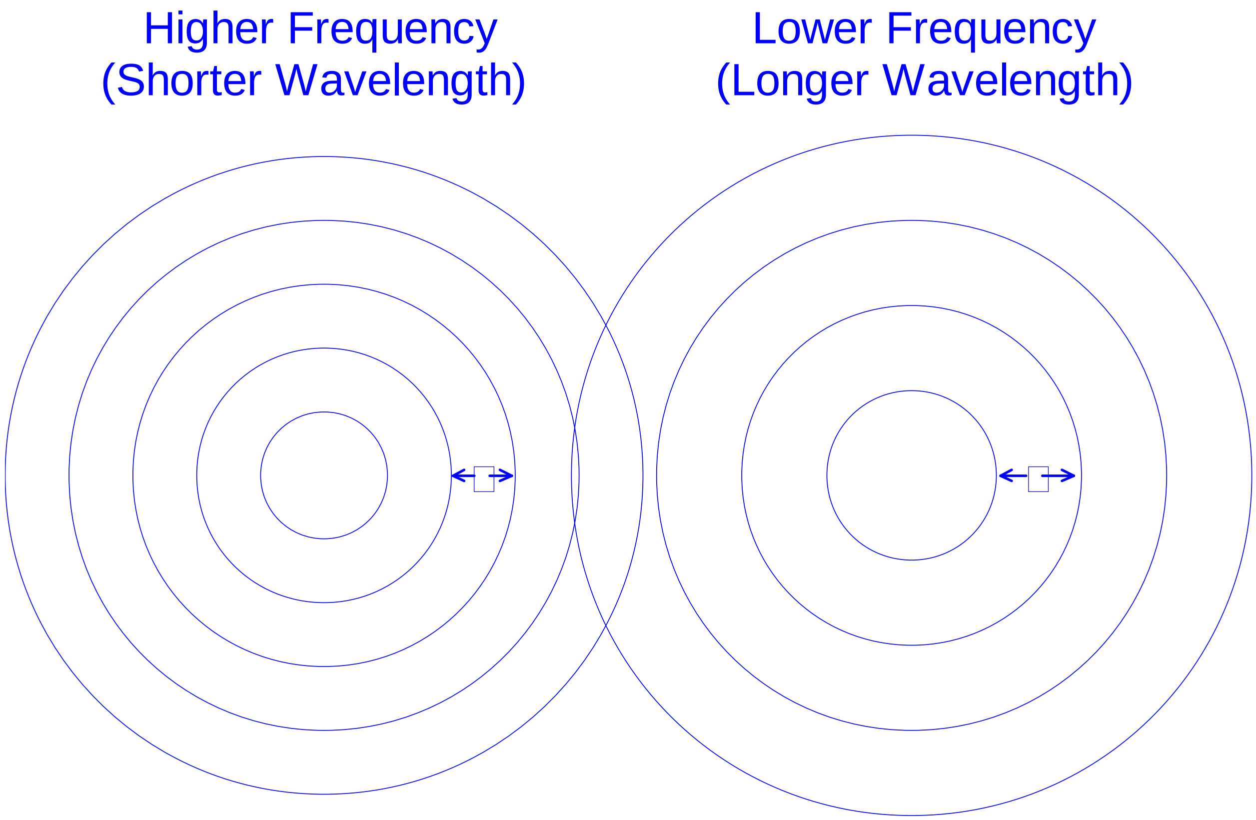


Figure 3-16. Wavelength is a measure of the distance between the crest of one cycle of a wave and the crest of the next cycle (or trough to trough or, in fact, the distance between any two corresponding points in the wave). Wavelength and frequency are related to one another. Because the wave has only a short time to travel from one cycle to the next, high frequencies produce short wavelengths. Conversely, because of the longer travel times, low frequencies produce long wavelengths.

circles (see Figure 3-16). Wavelength is simply the distance between the adjacent waves. Precisely the same concept can be applied to sound waves: wavelength is simply the distance between one compression wave and the next (or one rarefaction wave and the next or, more generally, the distance between any two corresponding points in adjacent waves). For our purposes, the most important point to be made about wavelength is that there is a simple relationship between frequency and wavelength. Using the puddle example, imagine that we begin by dipping our finger into the puddle at a very slow rate; that is, with a low "dipping frequency." Since the waves have a long period of time to travel from one dip to the next, the wavelength will be large. By the same reasoning, the wavelength becomes smaller as the "dipping frequency" is increased; that is, the time allowed for the wave to travel at high "dipping frequency" is small, so the wavelength is small. Wavelength is a measure of distance, and the formula for calculating wavelength is a straightforward algebraic rearrangement of the familiar "distance = rate · time" formula from junior high school.

$\lambda = c/f$, where:
- λ = wavelength
- c = the speed of sound
- f = frequency

By rearranging the formula, frequency can be calculated if wavelength and the speed of sound are known:

$f = c/\lambda$

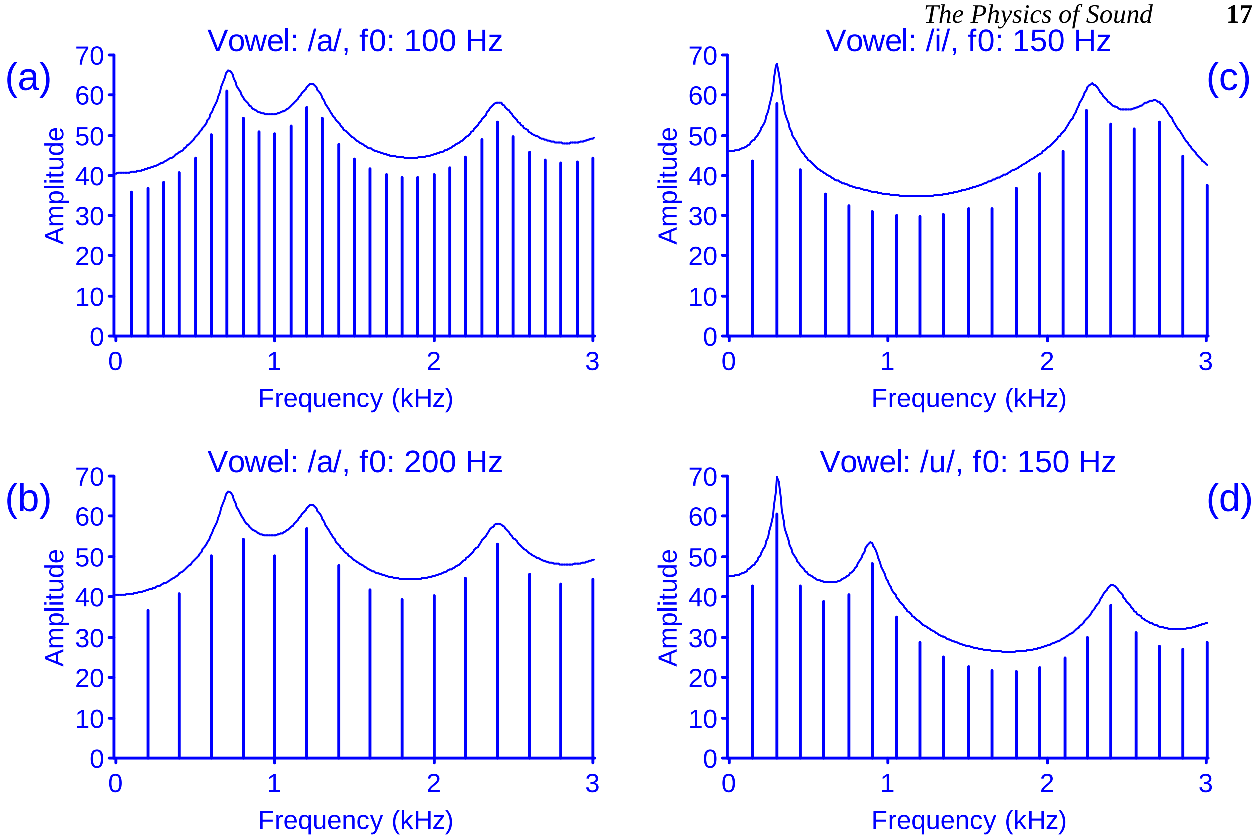


Figure 3-17. A spectrum envelope is an imaginary smooth line drawn to enclose an amplitude spectrum. Panels *a* and *b* show the spectra of two signals (the vowel /ɑ/) with different fundamental frequencies (note the differences in harmonic spacing) but very similar spectrum envelopes. Panels *c* and *d* show the spectra of two signals with different spectrum envelopes (the vowels /i/ and /u/ in this case) but the same fundamental frequencies (i.e., the same harmonic spacing).

Spectrum Envelope: The term **spectrum envelope** refers to an imaginary smooth line drawn to enclose an amplitude spectrum. Figure 3-17 shows several examples. This is a rather simple concept that will play a very important role in understanding certain aspects of auditory perception. For example, we will see that our perception of a perceptual attribute called **timbre** (also called **sound quality**) is controlled primarily by the shape of the spectrum envelope, and not by the fine details of the amplitude spectrum. The examples in Figure 3-17 show how differences in spectrum envelope play a role in signaling differences in one specific example of timbre called **vowel quality** (i.e., whether a vowel sounds like /i/ vs. /a/ vs. /u/, etc.). For example, panels *a* and *b* in Figure 3-17 show the vowel /å/ produced at two different fundamental frequencies. (We know that the fundamental frequencies are different because one spectrum shows wide harmonic spacing and the other shows narrow harmonic spacing.) The fact that the two vowels are heard as /a/ despite the difference in fundamental frequency can be attributed to the fact that these two signals have similar spectrum envelopes. Panels *c* and *d* in Figure 3-17 show the spectra of two signals with different spectrum envelopes but the same fundamental frequency (i.e., with the same harmonic spacing). As we will see in the chapter on auditory perception, differences in fundamental frequency are perceived as differences in pitch. So, for signals (a) and (b) in Figure 3-17, the listener will hear the same vowel produced at two different pitches. Conversely, for signals (c) and (d) in Figure 3-17, the listener will hear two different vowels produced at the same pitch. We will return to the concept of spectrum envelope in the chapter on auditory perception.

Amplitude Envelope: The term amplitude envelope refers to an imaginary smooth line that is drawn on top of a time domain signal. Figure 3-18 shows sinusoids that are identical except for their amplitude envelopes. It can be seen that the different amplitude envelopes reflect differences in the way the sounds are turned on and off. For example, panel *a* shows a signal that is turned on abruptly and turned off abruptly; panel *b* shows a signal that is turned on gradually and turned off abruptly; and so on. Differences in amplitude envelope have an important effect on the quality of a sound. As we will see in the chapter on auditory perception, amplitude envelope, along with spectrum envelope discussed above, is another physical parameter that affects **timbre** or **sound quality**. For

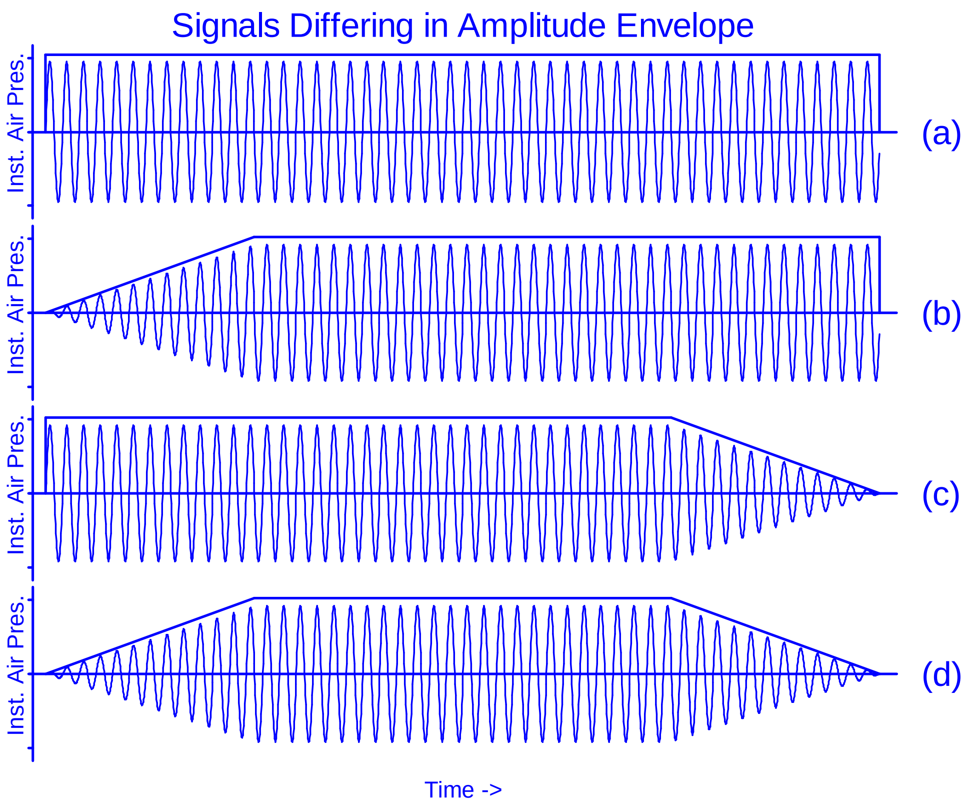


Figure 3-18. Amplitude envelope is an imaginary smooth line drawn to enclose a time-domain signal. This feature describes how a sound is turned on and turned off; for example, whether the sound is turned on abruptly and turned off abruptly (panel *a*), turned on gradually and turned off abruptly (panel *b*), turned on abruptly and turned off gradually (panel *c*), or turned on and off gradually (panel *d*).

example, piano players know that a given note will sound different depending on whether or not the damping pedal is used. Similarly, notes played on a stringed instrument such as a violin or cello will sound different depending on whether the note is plucked or bowed. In both cases, the underlying acoustic difference is amplitude envelope.

Acoustic Filters

As will be seen in subsequent chapters, acoustic filtering plays a central role in the processing of sound by the inner ear. The human vocal tract also serves as an acoustic filter that modifies and shapes the sounds that are created by the larynx and other articulators. For this reason, it is quite important to understand how acoustic filters work. In the most general sense, the term filter refers to a device or system that is selective about the kinds of things that are allowed to pass through versus the kinds of things that are blocked. An oil filter, for example, is designed to allow oil to pass through while blocking particles of dirt. Of special interest to speech and hearing science are **frequency selective** filters. These are devices that allow some frequencies to pass through while blocking or **attenuating** other frequencies. (The term **attenuate** means to weaken or reduce in amplitude).

A simple example of a frequency selective filter from the world of optics is a pair of tinted sunglasses. A piece of white paper that is viewed through red tinted sunglasses will appear red. Since the original piece of paper is white, and since we know that white light consists of all of the visible optical frequencies mixed in equal amounts, the reason that the paper appears red through the red tinted glasses is that optical frequencies other than those corresponding to red are being blocked or attenuated by the optical filter. As a result, it is primarily the red light that is being allowed to pass through. (Starting at the lowest optical frequency and going to the highest, light will appear red, orange, yellow, green, blue, indigo, and violet.)

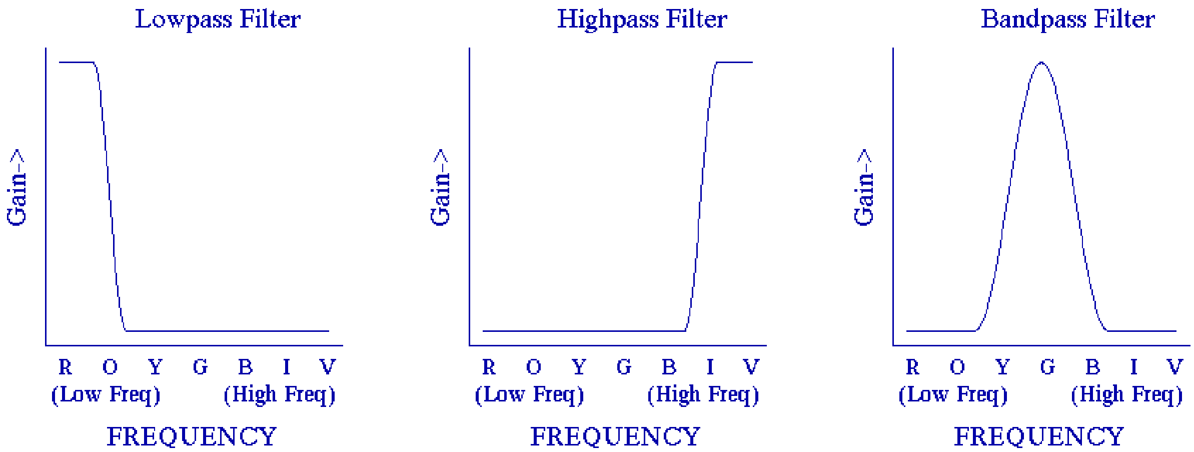


Figure 3-19. Frequency response curves for three optical filters. The lowpass filter on the left allows low frequencies to pass through, while attenuating or blocking optical energy at higher frequencies. The highpass filter in the middle has the opposite effect, allowing high frequencies to pass through, while attenuating or blocking optical energy at lower frequencies. The bandpass filter on the right allows a band of optical frequencies in the center of the spectrum to pass through, while attenuating or blocking energy at higher and lower frequencies.

A graph called a **frequency response curve** is used to describe how a frequency selective filter will behave. A frequency response curve is a graph showing how energy at different frequencies will be affected by the filter.

Specifically, a frequency response curve plots a variable called "gain" as a function of variations in the frequency of the input signal. Gain is the amount of amplification provided by the filter at different signal frequencies. Gains are interpreted as amplitude multipliers; for example, suppose that the gain of a filter at 100 Hz is 1.3. If a 100 Hz sinusoid enters the filter measuring 10 uPa, the amplitude at the output of the filter at 100 Hz will measure 13 µPa (10 µPa x 1.3 = 13 µPa). The only catch in this scheme is that gains can and very frequently are less than 1, meaning that the effect of the filter will be to attenuate the signal. For example, if the gain at 100 Hz is 0.5, a 10 µPa input signal at 100 Hz will measure 5 µPa at the output of the filter. When the filter gain is 1.0, the signal is unaffected by the filter; i.e., a 10 µPa input signal will measure 10 µPa at the output of the filter.

Figure 3-19 shows frequency response curves for several optical filters. Panel a shows a frequency response curve for the red optical filter discussed in the example above. If we put white light into the filter in panel a, the signal amplitude at the output of the filter will be high only when the frequency of the input signal is low. This is because the gain of the filter is high only in the low-frequency portion of the frequency-response curve. This is an example of a **lowpass** filter; that is, a filter that allows low frequencies to pass through. Panel b shows an optical filter that has precisely the reverse effect on an input signal; that is, this filter will allow high frequencies to pass through while attenuating low- and mid-frequency signals. A white surface viewed through this filter would therefore appear violet. This is an example of a **highpass** filter. Panel c shows the frequency response curve for a filter that allows a band of energy in the center of the spectrum to pass through while attenuating signal components of higher and lower frequency. A white surface viewed through this filter would appear green. This is called a **bandpass** filter.

Acoustic filters do for sound exactly what optical filters do for light; that is, they allow some frequencies to pass through while attenuating other frequencies. To get a better idea of how a frequency response curve is measured, imagine that we ask a singer to attempt to shatter a crystal wine glass with a voice signal alone. To see how the frequency response curve is created we have to make two rather unrealistic assumptions: (1) we need to assume that the singer is able to produce a series of *pure tones* of various frequencies (the larynx, in fact, produces a complex periodic sound and not a sinusoid), and (2) the amplitudes of these pure tones are always exactly the same. The wine glass will serve as the filter whose frequency response curve we wish to measure. As shown in Figure 3-20, we attach a vibration meter to the wine glass, and the reading on this meter will serve as our measure of output

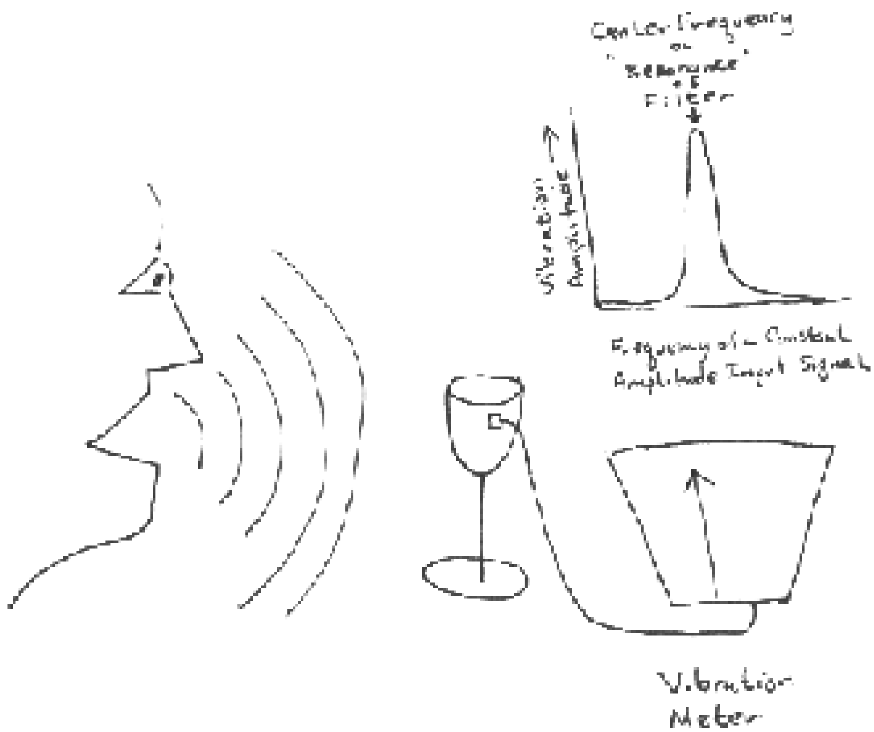


Figure 3-20. Illustration of how the frequency response curve of a crystal wine glass might be measured. Our singer produces a series of sinusoids that are identical in amplitude but cover a wide range of frequencies. (This part of the example is unrealistic: the human larynx produces a complex sound rather than a sinusoid.) The gain of the wine glass filter can be traced out by measuring the amplitude of vibration at the different signal frequencies.)

amplitude for the filter. For the purpose of this example, will assume that the signal frequency needed to break the glass is 500 Hz. We now ask the singer to produce a low frequency signal, say 50 Hz. Since this frequency is quite remote from the 500 Hz needed to break the glass, the output amplitude measured by the vibration meter will be

quite low. As the singer gets closer and closer to the required 500 Hz, the measured output amplitude will increase systematically until the glass finally breaks. If we assume that the glass does not break but rather reaches a maximum amplitude just short of that required to shatter the glass, we can continue our measurement of the frequency response curve by asking the singer to produce signals that are increasingly high in frequency. We would

find that the output amplitude would become lower and lower the further we got from the 500 Hz natural vibrating frequency of the wine glass. The pattern that is traced by our measures of output amplitude at each signal frequency would resemble the frequency response curve we saw earlier for green sunglasses; that is, we would see the frequency response curve for a bandpass filter.

Additional Comments on Filters

Cutoff Frequency, Center Frequency, Bandwidth. The top panel of Figure 3-21 shows frequency response curves for two lowpass filters that differ in a parameter called **cutoff frequency**. Both filters allow low frequencies to pass through while attenuating high frequencies; the filters differ only in the frequency at which the attenuation begins. The bottom panel of Figure 3-21 shows two highpass filters that differ in cutoff frequency. There are two additional terms that apply only to bandpass filters. In our wineglass example above, the natural vibrating frequency of the wine glass was 300 Hz. For this reason, when the frequency response curve is measured, we find that the wine glass reaches its maximum output amplitude at 300 Hz. This is called the **center frequency** or **resonance** of the filter. It is possible for two bandpass filters to have the same center frequency but differ with respect to a property called

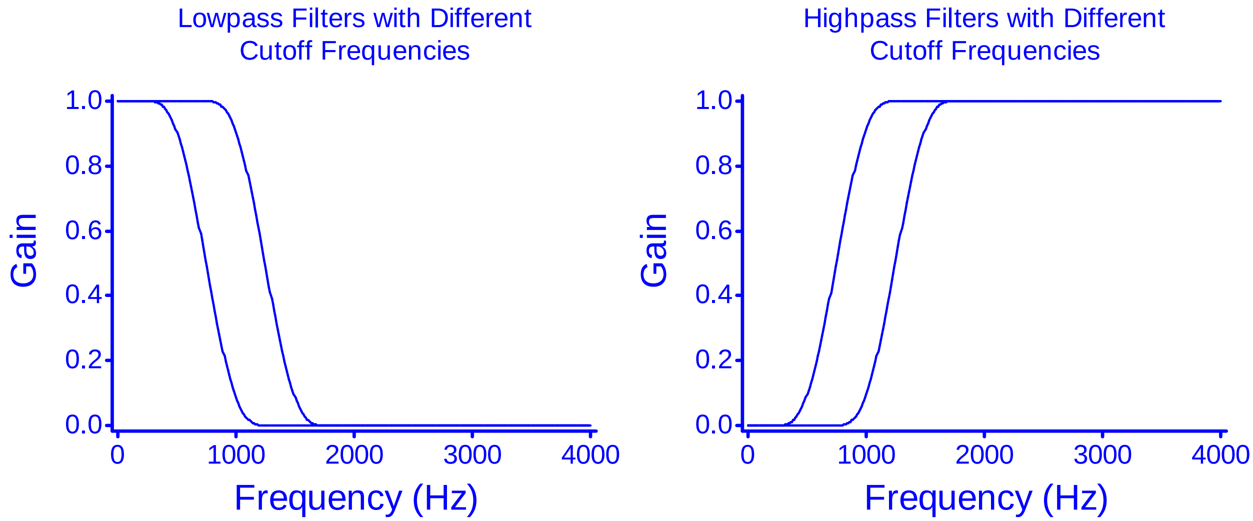


Figure 3-21. Lowpass and highpass filters differing in cutoff frequency.

bandwidth. Figure 3-22 shows two filters that differ in bandwidth. The tall, thin frequency response curve describes a **narrow band** filter. For this type of filter, output amplitude reaches a very sharp peak at the center frequency and drops off abruptly on either side of the peak. The other frequency response curve describes a ***wide band*** filter (also called ***broad band***). For the wide band filter, the peak that occurs at the resonance of the filter is less sharp and the drop in output amplitude on either side of the center frequency is more gradual.

Fixed vs. Variable Filters. A fixed filter is a filter whose frequency response curve cannot be altered. For example, an engineer might design a lowpass filter that attenuates at frequencies above 500 Hz, or a bandpass filter that passes with a center frequency of 1,000 Hz. It is also possible to create a filter whose characteristics can be varied. For example, the tuning dial on a radio controls the center frequency of a narrow bandpass filter that allows a single radio channel to pass through while blocking channels at all other frequencies. The human vocal tract is an example

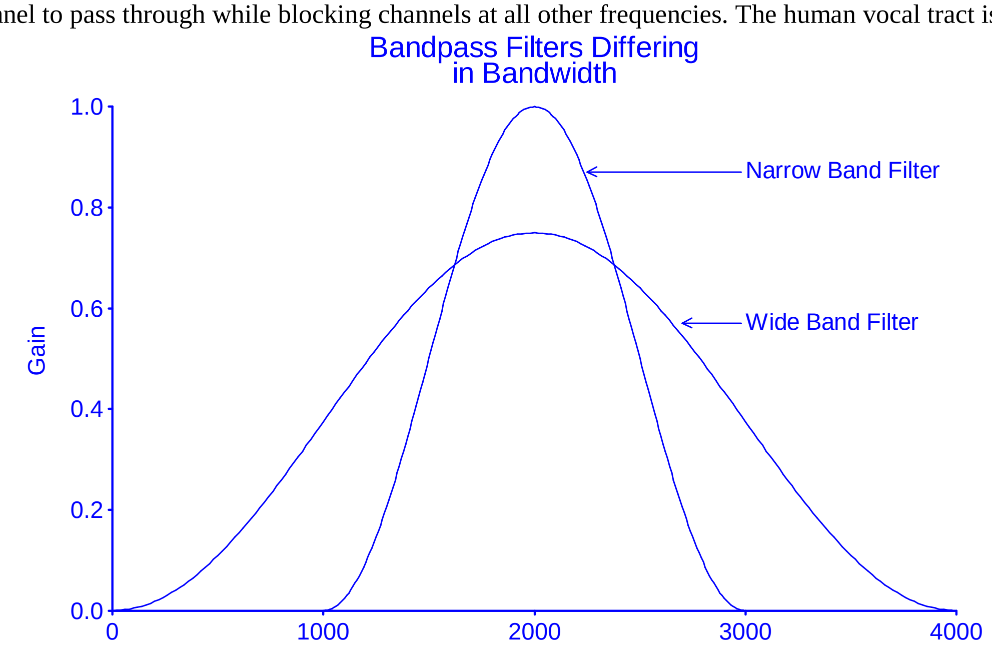


Figure 3-22. Frequency response curves for two bandpass filters with identical center frequencies but different bandwidths. Both filters pass a band of energy centered around 2000 Hz, but the narrow band filter is more selective than the wide band filter; that is, gain decreases at a higher rate above and below the center frequency for the narrow band filter than for the wide band filter

of a variable filter of the most spectacular sort. For example: (1) during the occlusion interval that occurs in the production of a sound like /b/, the vocal tract behaves like a lowpass filter; (2) in the articulatory posture for sounds like /s/ and /sh/ the vocal tract behaves like a highpass filter; and (3) in the production of vowels, the vocal tract behaves like a series of bandpass filters connected to one another, and the center frequencies of these filters can be adjusted by changing the positions of the tongue, lips, and jaw. To a very great extent, the production of speech involves making adjustments to the articulators that have the effect of setting the vocal tract filter in differ modes to produce the desired sound quality. We will have much more to say about this in later chapters.

Frequency Response Curves vs. Amplitude Spectra. It is not uncommon for students to confuse a frequency response curve with an amplitude spectrum. The axis labels are rather similar: an amplitude spectrum plots amplitude on the y axis and frequency on the x axis, while a frequency response curve plots gain on the y axis and frequency on the x axis. The apparent similarities are deceiving, however, since a frequency response curve and an amplitude spectrum display very different kinds of information. The difference is that *an amplitude spectrum describes a sound while a frequency response curve describes a filter*. For any given sound wave, an amplitude spectrum tells us what frequencies are present with what amplitudes. A frequency response curve, on the other hand, describes a filter, and for that filter, it tells us what frequencies will be allowed to pass through and what frequencies will be attenuated. Keeping these two ideas separate will be quite important for understanding the key role played by filters in both hearing and speech science.

Resonance

The concept of **resonance** has been alluded to on several occasions but has not been formally defined. The term resonance is used in two different but very closely related ways. The term resonance refers to: (1) the phenomenon of **forced vibration**, and (2) **natural vibrating frequency** (also **resonant frequency** or **resonance frequency**) To gain an appreciation for both uses of this term, imagine the following experiment. We begin with two identical tuning forks, each tuned to 435 Hz. Tuning fork A is set into vibration and placed one centimeter from tuning fork B, but not touching it. If we now hold tuning fork B to a healthy ear, we will find that it is producing a 435 Hz tone that is faint but quite audible, despite the fact that it was not struck and did not come into physical contact with tuning fork A. The explanation for this "action-at-a-distance" phenomenon is that the sound wave generated by tuning fork A forces tuning fork B into vibration; that is, the series of compression and rarefaction waves will alternately push and pull the tuning fork, resulting in vibration at the frequency being generated by tuning fork A. The phenomenon of forced vibration is not restricted to this "action-at-a-distance" case. The same effect can be demonstrated by placing a vibrating tuning fork in contact with a desk or some other hard surface. The intensity of the signal will increase dramatically because the tuning fork is forcing the desk to vibrate, resulting in a larger volume of air being compressed and rarefied.[3]

Returning to our original tuning fork experiment, suppose that we repeat this test using two *mismatched* tuning forks; for example, tuning fork A with a natural frequency of 256 Hz and tuning fork B with a natural vibrating frequency of 435 Hz. If we repeat the experiment – setting tuning fork A into vibration and holding it one centimeter from tuning fork B – we will find that tuning fork B does not produce an audible tone. The reason is that forced vibration is most efficient when the frequency of the driving force is closest to the natural vibration frequency of the object that is being forced to vibrate. Another way to think about this is that tuning fork B in these experiments is behaving like a filter that is being driven by the signal produced by tuning fork A. Tuning forks, in fact, behave like rather narrow bandpass filters. In the experiment with matched tuning forks, the filter was being driven by a signal frequency corresponding to the peak in the filter's frequency response curve. Consequently, the filter produced a great deal of energy at its output. In the experiment with mismatched tuning forks, the filter is being driven by a signal that is remote from the peak in the filter's frequency response curve, producing a low amplitude output signal.

To summarize, resonance refers to the ability of one vibrating system to force another system into vibration. Further, the amplitude of this forced vibration will be greater as the frequency of the driving force approaches the natural vibrating frequency (resonance) of the system that is being forced into vibration.

---

[3]The increase in intensity that would occur as the tuning fork is placed in contact with a hard surface does not mean that additional energy is created. The increase in intensity would be offset by a decrease in the duration of the tone, so the total amount of energy would not increase relative to a freely vibrating tuning fork.

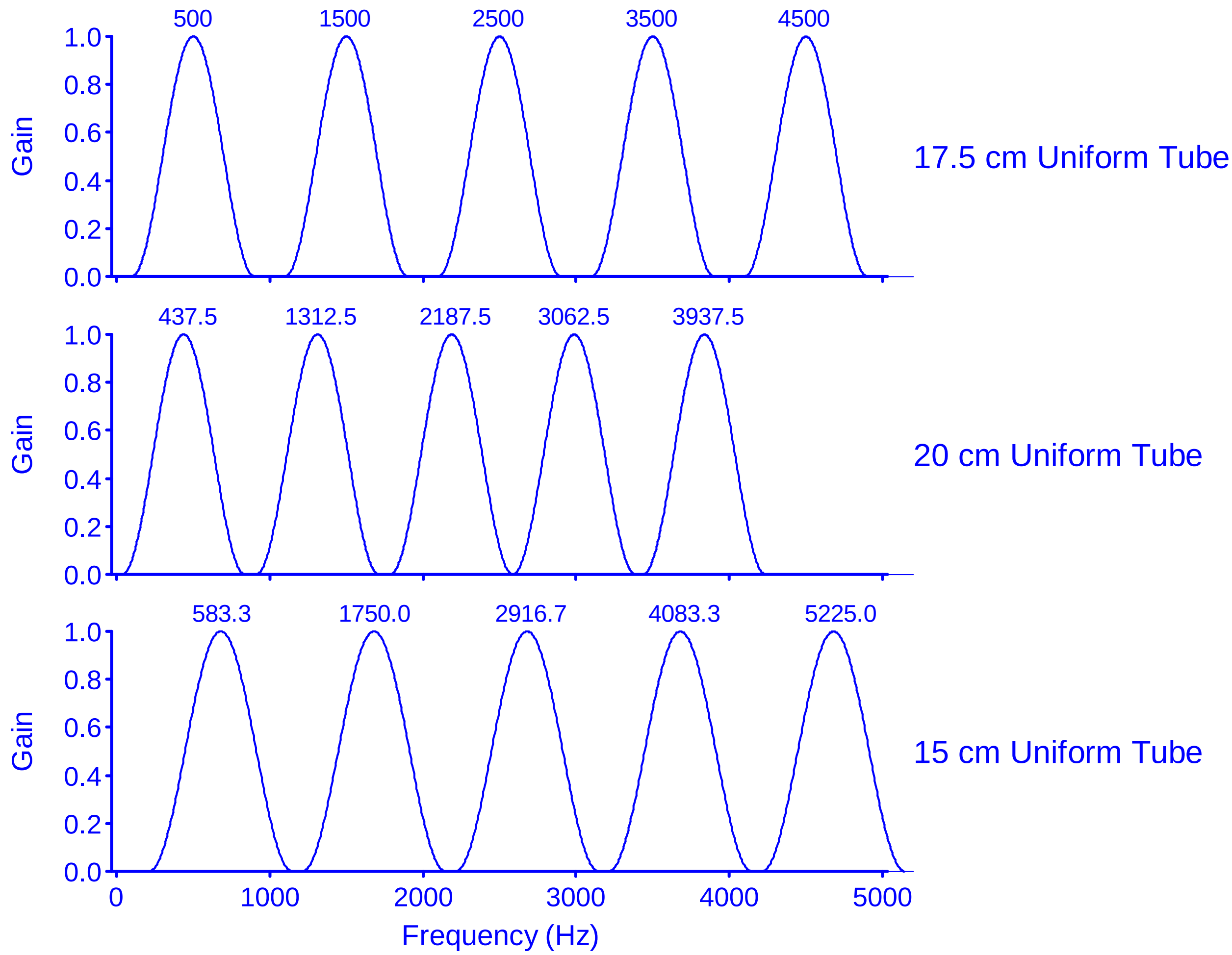


Figure 3-23. Frequency response curves for three uniform tubes open at one end and closed at the other. These kinds of tubes have an infinite number of resonances at odd multiples of the lowest resonance. As the figure shows, shortening the tube shifts all resonances to higher frequencies while lengthening the tube shifts all resonances to lower frequencies.

Cavity Resonators

An air-filled cavity exhibits frequency selective properties and should be considered a filter in precisely the way that the tuning forks and wine glasses mentioned above are filters. The human vocal tract is an air-filled cavity that behaves like a filter whose frequency response curve varies depending on the positions of the articulators. Tuning forks and other simple filters have a single resonant frequency. (Note that we will be using the terms "natural vibrating frequency" and "resonant frequency" interchangeably.) Cavity resonators, on the other hand, can have an infinite number of resonant frequencies.

A simple but very important cavity resonator is the **uniform tube.** This is a tube whose cross-sectional area is the same (uniform) at all points along its length. A simple water glass is an example of a uniform tube. The method for determining the resonant frequency pattern for a uniform tube will vary depending on whether the tube is closed at both ends, open at both ends, or closed at just one end. The configuration that is most directly applicable to problems in speech and hearing is the uniform tube that is closed at one end and open at the other end. The ear canal, for example, is approximately uniform in cross-sectional area and is closed medially by the ear drum and open laterally. Also, in certain configurations the vocal tract is approximately uniform in cross-sectional area and is effectively closed from below by the vocal folds and open at the lips. The resonant frequencies for a uniform tube closed at one end are determined by its length. The lowest resonant frequency (F_1) for this kind of tube is given by:

$F_1 = c/4L$, where: c = the speed of sound
L = the length of the tube

For example, for a 17.5 cm tube, $F_1 = c/4L = 35000/70 = 500$ Hz. This tube will also have an infinite number of higher frequency resonances at *odd* multiples of the lowest resonance:

$F_1 = F_1 \cdot 1 = 500$ Hz
$F_2 = F_1 \cdot 3 = 1{,}500$ Hz
$F_3 = F_1 \cdot 5 = 2{,}500$ Hz
$F_4 = F_1 \cdot 7 = 3{,}500$ Hz

The frequency response curve for this tube for frequencies below 4000 Hz is shown in the solid curve in Figure 3-23. Notice that the frequency response curve shows peaks at 500, 1500, 2500, and 3500 Hz, and valleys in between these peaks. The frequency response curve, in fact, looks like a number of bandpass filters connected in series with one another. It is important to appreciate that what we have calculated here is a series of natural vibrating frequencies of a tube. What this means is that the tube will respond best to forced vibration if the tube is driven by signals with frequencies at or near 500 Hz, 1500 Hz, 2500 Hz, and so on. Also, the *resonant* frequencies that were just calculated should not be confused with *harmonics*. Harmonics are frequency components that are present in the amplitude spectra of complex periodic sounds; resonant frequencies are peaks in the frequency response curve of filters.

We next need to see what will happen to the resonant frequency pattern of the tube when the tube length changes. If the tube is lengthened to 20 cm:

$F_1 = c/4L = 35{,}000/80 = 437.5$ Hz
$F_2 = F_1 \cdot 3 = 1{,}312.5$ Hz
$F_3 = F_1 \cdot 5 = 2{,}187.5$ Hz
$F_4 = F_1 \cdot 7 = 3{,}062.5$ Hz

It can be seen that lengthening the tube from 17.5 cm to 20 cm has the effect of shifting all of the resonant frequencies downward (see Figure 3-23). Similarly, shortening the tube has the effect of shifting all of the resonant frequencies upward. For example, the resonant frequency pattern for a 15 cm tube would be:

$F_1 = c/4L = 35{,}000/60 = 583.3$ Hz
$F_2 = F_1 \cdot 3 = 1{,}750$ Hz
$F_3 = F_1 \cdot 5 = 2{,}916.7$ Hz
$F_4 = F_1 \cdot 7 = 4{,}083.3$ Hz

The general rule is quite simple: all else being equal, long tubes have low resonant frequencies and short tubes have high resonant frequencies. This can be demonstrated easily by blowing into bottles of various lengths. The longer bottles will produce lower tones than shorter bottles. This effect is also demonstrated every time a water glass is filled. The increase in the frequency of the sound that is produced as the glass is filled occurs because the resonating cavity becomes shorter and shorter as more air is displaced by water. This simple rule will be quite useful. For example, it can be applied directly to the differences that are observed in the acoustic properties of speech produced by men, women, and children, who have vocal tracts that are quite different in length.

Resonant Frequencies and Formant Frequencies

The term "resonant frequency" refers to natural vibrating frequency or, equivalently, to a peak in a frequency response curve. For reasons that are entirely historical, if the filter that is being described happens to be a human vocal tract, the term **formant frequency** is generally used. So, one typically refers to the *formant frequencies* of the vocal tract but to the *resonant frequencies* of a plastic tube, the body of a guitar, the diaphragm of a loudspeaker, or most any other type of filter other than the vocal tract. This is unfortunate since it is possible to get the mistaken idea that formant frequencies and resonant frequencies are different sorts of things. The two terms are, in fact, fully synonymous.

The Decibel Scale

The final topic that we need to address in this chapter is the representation of signal amplitude using the decibel scale. The decibel scale is a powerful and immensely flexible scale for representing the amplitude of a sound wave. The scale can sometimes cause students difficulty because it differs from most other measurement scales in not just one but two ways. Most of the measurement scales with which we are familiar are **absolute** and **linear**. The decibel scale, however, is **relative** rather than absolute, and **logarithmic** rather than linear. Neither of these characteristics is terribly complicated, but in combination they can make the decibel scale appear far more obscure than it is. We will examine these features one at a time, and then see how they are put together in building the decibel scale.

Linear vs. Logarithmic Measurement Scales

Most measurement scales are linear. To say that a measurement scale is linear means that it is based on equal *additive* distances. This is such a common feature of measurement scales that we do not give it much thought. For example, on a centigrade (or Fahrenheit) scale for measuring temperature, going from a temperature of 90^o to a temperature of 91^o involves *adding* one 1^o. One rather obvious consequence of this simple additivity rule is that the difference in temperature between 10^o and 11^o is the same as the difference in temperature between 90^o and 91^o. However, there are scales for which this additivity rule does not apply. One of the best known examples is the Richter scale that is used for measuring seismic intensity. The difference in seismic intensity between Richter values of 4.0 and 5.0, 5.0 and 6.0, 6.0 and 7.0 is not some constant amount of seismic intensity, but rather a constant *multiple*. Specifically, a 7.0 on the Richter scale indicates an earthquake that is 10 times greater in intensity than an earthquake that measures 6.0 on the Richter scale. Similarly, an 8.0 on the Richter scale is 10 times greater in intensity than a 7.0. Whenever jumping from one scale value to the next involves multiplying by a constant rather than adding a constant, the scale is called logarithmic. (The multiplicative constant need not be 10. See Box 3-2 for an example of a logarithmic scale – an octave progression – that uses 2 as the constant.) Another way of making the same point is to note that the values along the Richter scale are exponents rather than ordinary numbers; for example, a Richter value of 6 indicates a seismic intensity of 10^6, a Richter value of 7 indicates a seismic intensity of 10^7, etc. The Richter values can, of course, just as well be referred to as *powers* or *logarithms* since both of these terms are synonyms for *exponent*. The decibel scale is an example of a logarithmic scale, meaning that it is based on equal multiples rather than equal additive distances.

Absolute vs. Relative Measurement Scales

A simple example of a relative measurement scale is the **Mach** scale that is used by rocket scientists to measure speed. The Mach scale measures speed not in absolute terms but in relation to the speed of sound. For example, a missile at Mach 2.0 is traveling at twice the speed of sound, while a missile at Mach 0.9 is traveling at 90% of the speed of sound. So, the Mach scale does not represent a measured speed (S_m) in absolute terms, but rather, represents a measured speed in relation to a reference speed (S_m/S_r). The reference that is used for the Mach scale is the speed of sound, so a measured absolute speed can be converted to a relative speed on the Mach scale by simple division. For example, taking 783 mph as the speed of sound, 1,200 mph = 1200/783 = Mach 1.53. The decibel scale also exploits this relative measurement scheme. The decibel scale does not represent a measured intensity (I_m) in absolute terms, but rather, represents the ratio of a measured intensity to a reference intensity (I_m/I_r).

The decibel scale is trickier than the Mach scale in one important respect. For the Mach scale, the reference is always the speed of sound, but for the decibel scale, many different references can be used. In explaining how the decibel scale works, we will begin with the commonly used intensity reference of 10^{-12} w/m^2 (watts per square meter), which is approximately the intensity that is required for an average normal hearing listener to barely detect a 1,000 Hz pure tone. So, for our initial pass through the decibel scale, 10^{-12} w/m^2 will serve as I_r, and will perform the same function that the speed of sound does for the Mach scale. Table 3-1 lists several sounds that cover a very broad range of intensities. The second column shows the measured intensities of those sounds, and the third column shows the ratio of those intensities to our reference intensity. Whispered speech, for example, measures approximately 10^{-8}

w/m^2, which is 10,000 times more intense than the reference intensity ($10^{-8}/10^{-12} = 10^4 = 10{,}000$). The main point to be made about column 3 is that the ratios become very large very soon. Even a moderately intense sound like conversational speech is 1,000,000 times more intense than the reference intensity. The awkwardness of dealing with these very large ratios has a very simple solution. Column 4 shows the ratios written in exponential notation, and column 5 simplifies the situation even further by recording the exponent only. The term exponent and the term logarithm are synonymous, so the measurement scheme that is expressed by the numbers in column 5 can be summarized as follows: (1) divide a measured intensity by a reference intensity (in this case, 10^{-12} w/m^2), (2) take the logarithm of this ratio (i.e., write the number in exponential notation and keep the exponent only). This method, in fact, is a completely legitimate way to represent signal intensity. The unit of measure is called the bel, after A.G. Bell, and the formula is:

bel = $\log_{10} I_m/I_r$, where: I_m = a measured intensity
I_r = a reference intensity

Table 3-1. Sound intensities and intensity ratios showing how the decibel scale is created. Column 2 shows the measured intensities (I_m) of several sounds. Column 3 shows the ratio of these intensities to a reference intensity of 10^{-12} w/m^2. Column 4 shows the ratio written in exponential notation while column 5 shows the exponent only. The last column shows the intensity ratio expressed in decibels, which is simply the logarithm of the intensity ratio multiplied by 10.

Sound	**Measured Intensity (I_m)**	**Ratio (I_m/I_r)**	**Ratio in Exp. Not.**	**Exponent ($\log_{10}$)**	**Decibel ($10 \times \log_{10}$)**
Threshold @ 1 kHz	10^{-12} w/m^2	1	10^0	0	0
Whisper	10^{-8} w/m^2	10,000	10^4	4	40
Conversational Speech	10^{-6} w/m^2	1,000,000	10^6	6	60
City Traffic	10^{-4} w/m^2	100,000,000	10^8	8	80
Rock & Roll	10^{-2} w/m^2	10,000,000,000	10^{10}	10	100
Jet Engine	10^0 w/m^2	1,000,000,000,000	10^{12}	12	120

Legitimate or not, the bel finds its sole application in textbooks attempting to explain the decibel. For reasons that are purely historical, the $\log_{10}$ of the intensity ratio is multiplied by 10, changing bel into the decibel (dB). As shown in the last column of Table 3-1, this has the very simple effect of turning 4 bels into 40 decibels, 8 bels into 80 decibels, etc. The formula for the decibel, then, is:

$dB_{IL} = 10 \log_{10} I_m/I_r$, where:

I_m = a measured intensity
I_r = a reference intensity

The designation "IL" stands for **intensity level**, and it indicates that the underlying measurements are of sound *intensity* and not sound *pressure*. As will be seen below, a different version of this formula is needed if sound pressure measurements are used. The multiplication by 10 in the dB_{IL} formula is a simple operation, but it can sometimes have the unfortunate effect of making the formula appear more obscure that it is. The decibel values that are calculated, however, should be readily interpretable. For example, 30 dB_{IL} means 3 factors of 10 more intense than I_r, 60 dB_{IL} means 6 factors of 10 more intense than I_r, and 90 dB_{IL} means 9 factors of 10 more intense than I_r.

Deriving a Pressure Version of the dB Formula

In a simple world, we would be finished with the decibel scale. The problem is that the formula is based on measurements of sound intensity, but as a purely practical matter sound intensity is difficult to measure. Sound pressure, on the other hand, is quite easy to measure. An ordinary microphone, for example, is a pressure sensitive device. The problem, then, is that the decibel is defined in terms of intensity measurements, but the measurements that are actually used will nearly always be measures of sound pressure. This problem can be addressed since there is a predictable relationship between intensity (I) and pressure (E): intensity is proportional to pressure squared:

$$I \propto E^2$$

Knowing this relationship allows us to create a completely equivalent version of the decibel formula that will work when sound pressure measurements are used instead of sound intensity measurements. All we need to do is substitute *squared* pressure measurements in place of the intensity measurements:

$dB_{IL} = 10 \log_{10} I_m/I_r$ (intensity version of formula)
$dB_{SPL} = 10 \log_{10} E^2_m/E^2$ (pressure version of formula)

The designation "SPL" stands for ***sound pressure level***, and it indicates that measures of sound pressure have been used and not measures of sound intensity. Although the dB_{SPL} formula shown here will work fine, it will almost never be seen in this form. The reason is that the formula is algebraically rearranged so that the squaring operation is not needed. The algebra is shown below:

(1) $dB_{IL} = 10 \log_{10} I_m/I_r$ (the intensity version of the formula)

(2) $dB_{SPL} = 10 \log_{10} E^2_m/E^2$ (measures of E^2 replace measures of I because $I \propto E^2$)

(3) $dB_{SPL} = 10 \log_{10} (E_m/E_r)^2$ ($a^2/b^2 = (a/b)^2$)

(4) $dB_{SPL} = 10 \cdot 2 \log_{10} E_m/E_r$ (this is the only tricky step: $\log a^b = b \log a$)

(5) $dB_{SPL} = 20 \log_{10} E_m/E_r$ ($2 \cdot 10 = 20$)

With the possible exception of the fourth step,[4] the algebra is straightforward, but the details of the derivation are less important than the following general points:

1. The decibel formula is defined in terms of intensity ratios. The basic formula is;

   $dB_{IL} = 10 \log_{10} I_m/I_r$.

2. While sound intensity is difficult to measure, sound pressure is easy to measure. It is therefore necessary to derive a version of the decibel formula that works when measures of sound pressure are used instead of sound intensity.

---

[4] Step 4 is the only tricky part of derivation. The reason it works is that squaring a number and then taking a log is the same as taking the log first, and then multiplying the log by 2. For example, note that the two calculations below produce the same result:

$\log_{10} 100^2 = \log_{10} 10{,}000 = 4$ (square first, then take the log)
$\log_{10} 100^2 = (\log_{10} 100) \times 2 = 2 \times 2 = 4$ (take the log, then multiply by 2)

3. The derivation of the pressure version of the formula is based *entirely* on the fact that intensity is proportional to pressure squared ($I \propto E^2$). This allows measures of E^2 to replace measures of I, turning: $dB_{IL} = 10 \log_{10} I_m/I_r$ into $dB_{SPL} = 10 \log_{10} E^2_m/E_r^2$. A few algebra tricks are applied to turn this formula into the more aesthetically pleasing final version: $dB_{SPL} = 20 \log_{10} E_m/E_r$.

4. The two versions of the formula are *fully equivalent* to one another (see Box 3-3).

This last point about the equivalence of the intensity and sound pressure versions of the formula is explained in some detail in Box 3-3, but the basic point is quite simple. The pressure version of the dB formula was derived from the intensity version of the formula through algebraic manipulations (based on this relationship: $I \propto E^2$). The whole

---

Box 3-2

HARMONICS, OCTAVES, LINEAR SCALES, AND LOGARITHMIC SCALES

As we will see when the decibel scale is introduced, there is an important distinction to be made between ***linear*** scales, which are quite common, and ***logarithmic*** scales, which are less common but quite important. This distinction can be illustrated by examining the difference between a harmonic progression and an octave progression. Notice that in a harmonic progression, the spacing between the harmonics is always the same; that is, the difference between H_1 and H_2 is the same as the difference between H_2 and H_3, and so on. This is because increases in frequency between one harmonic and the next involve *adding a constant*, with the constant being the fundamental frequency. For example:

H_1	500
H_2	1000 (add 500)
H_3	1500 (add 500)
H_4	2000 (add 500)
.	.
.	.
.	.

To get from one scale value to another on an octave progression involves multiplying by a constant rather than adding a constant. For example, an octave progression starting at 500 Hz looks like this:

O_1	500
O_2	1000 (multiply by 2)
O_3	2000 (multiply by 2)
O_4	4000 (multiply by 2)
	. .
	. .
	. .

As a result of the fact that we are multiplying by a constant rather than adding a constant, the spacing is no longer even (i.e., the spacing between O_1 and O_2 is 500 Hz, the spacing between O_2 and O_3 is 1000 Hz, and so on). The point to be made of this is that there are two fundamentally different kinds of scales: (1) scales like harmonic progressions that are created by adding a constant, which are by far the more common, and (2) scales like octave progressions that are created by multiplying by a constant. Scales that are created by adding a constant are called **linear** scales, while scales that are created by multiplying by a constant are called **logarithmic** scales. Note that for an octave progression, the multiplier happens to be 2, meaning that progressing from one frequency to an octave above that frequency involves multiplication by 2. However, a logarithmic scale can be built using any multiplier. We will return to the distinction between linear and logarithmic scales when we talk about the decibel scale, and there we will see that a logarithmic scale is built around multiplication by a constant value of 10 rather than 2.

---

point of algebra, of course, is to keep the expression on the left equal to the expression on the right. The simple and useful point that emerges from this is this: If an intensity meter shows that a given sound measures 60 dB_{IL}, for example, *a pressure meter will show that the same sound measures exactly 60* dB_{SPL}. (This may seem counterintuitive due to the differences in the formulas, but see Box 3-3 for the explanation.) The equivalence of the two versions of the dB formula greatly simplifies the interpretation of sound levels that are expressed in decibels.

References

The reference that is used for the Mach scale is always the speed of sound. One of the virtues of the decibel scale is that any reference can be used as long as it is clearly specified. The only reference that has been mentioned so far is 10^{-12} w/m^2, which is roughly the audibility threshold for a 1,000 Hz pure tone. This is a standard reference intensity, and unless otherwise stated it should be assumed that this is used when a signal level is reported in dB_{IL}. The standard reference that is used for dB_{SPL} is 20 μPa, so when a signal level is reported in dB_{SPL} it should be assumed that this reference is used unless otherwise stated.[5]

Many references besides these two standard references can be used. For example, suppose that a speech signal is presented to a listener at an average level of 3500 μPa in the presence of a noise signal whose average sound pressure is 1400 μPa. The speech-to-noise ratio (S/N) can be represented on a decibel scale, using the level of the speech as E_m and the level of the noise as E_r:

$$\begin{aligned} dB_{s/n} &= 20 \log_{10} E_m / E_r \\ &= 20 \log_{10} 3500/1400 \\ &= 20 \log_{10} 2.5 \\ &= 20\ (0.39794) \\ &= 7.96 \text{ dB} \end{aligned}$$

To take one more example, assume that a voice patient prior to treatment produces sustained vowels that average 2300 μPa. Following treatment the average sound pressures increase to 8890 μPa. The improvement in sound pressure (post-treatment relative to pre-treatment) can be represented on a decibel scale:

$$\begin{aligned} dB_{Improvement} &= 20 \log_{10} E_{post}/E_{pre} \\ &= 20 \log_{10} 8890/2300 \\ &= 20 \log_{10} (3.86522) \\ &= 20\ (0.58717) \\ &= 11.74 \text{ dB} \end{aligned}$$

A final example can be used to make the point that the decibel scale can be used to represent intensity ratios for any type of energy, not just sound. Bright sunlight has a luminance measuring 100,000 cd/m^2 (candela per square meter). Light from a barely visible star, on the other hand, has a luminance measuring 0.0001 cd/m^2. We can now ask how much more luminous bright sunlight is in relation to barely visible star light, and the dB scale can be used to represent this value. Since the underlying physical quanities here are measures of electromagnetic intensity, we want the intensity version of the formula rather than the pressure version.

$$\begin{aligned} dB &= 10 \log_{10} I_{sunlight}/I_{starlight} \\ &= 10 \log_{10} 100000/0.0001 \\ &= 10 \log_{10} 10^5/10^{-4} \\ &= 10 \log_{10} 10^9 \text{ (division is done by subtracting exponents: } 5 - (-4) = 9) \\ &= 10\ (9) \\ &= 90 \text{ dB} \end{aligned}$$

---

[5] The standard pressure reference for dB_{SPL} is sometimes given as 0.0002 dynes/cm^2 rather than 20 μPa. These two sound pressures are identical, however, in exactly the same sense that 4 quarts and 1 gallon are identical. Likewise, the standard reference for dB_{IL} is often given as 10^{-16} w/cm^2 instead of 10^{-12} w/m^2. These two intensities are also identical.

The fact that we are measuring light rather than sound makes no difference: a decibel is 10 $\log_{10} I_m/I_r$ (or, equivalently, 20 $\log_{10} E_m/E_r$), regardless of whether the energy comes from sound, light, electrical current, or any other type of energy.

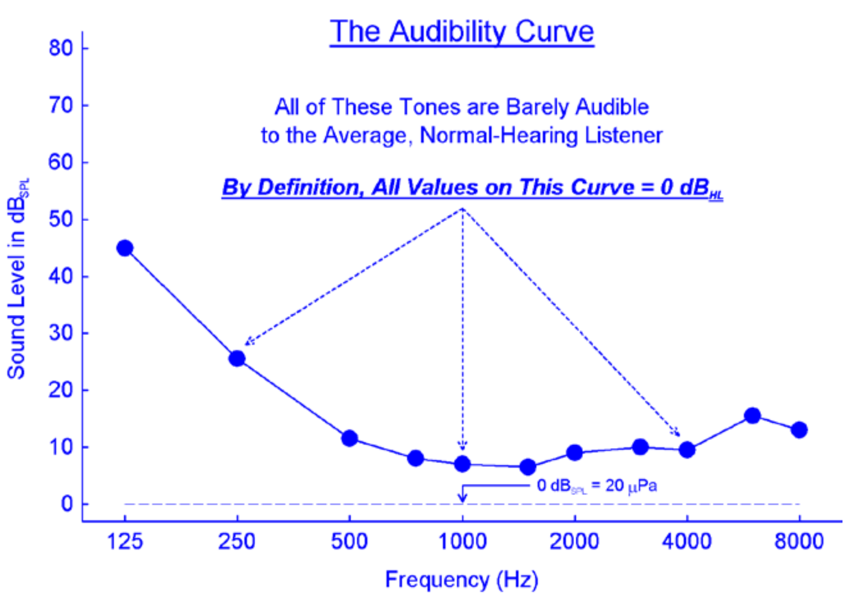
dB Hearing Level (dB_{HL})

The dB Hearing Level (dB_{HL}) scale was developed specifically for testing hearing sensitivity for pure tones of different frequencies. The sound-level dials on clinical audiometers,[6] for example, are calibrated in dB_{HL} rather than dB_{SPL}. To understand the motivation for the dB_{HL} scale examine Figure 3-24, which shows the sound level (in dB_{SPL}) required for the average, normal-hearing listener to barely detect pure tones at frequencies between 125 and 8000 Hz. This is called the **audibility curve** and the simple but very important point to notice about this graph is that the curve is not a flat line; that is, the ear is clearly more sensitive at some frequencies than others. The differences in sensitivity are quite large in some cases. For example, the average normal-hearing listener will barely detect a 1000 Hz pure tone at 7 dB_{SPL}, but at 125 Hz the sound level needs to be cranked all the way up to 45 dB_{SPL}, an increase in intensity of nearly 4000:1. Now suppose we were to test pure-tone sensitivity using an audiometer that is calibrated in dB_{SPL}. Imagine that a listener barely detects a 1000 Hz pure tone at 25 dB_{SPL}. Does this listener have a hearing loss, and if so how large? The only way to answer this question is to consult the data in Figure 3-24, which shows that the threshold of audibility for the average normal hearing listener at 1000 Hz is 7 dB_{SPL}. This means that the hypothetical listener in this example has a hearing loss of 25-7 = 18 dB. Suppose further that the same listener detects a 250 Hz tone at 20 $dB_{SPL.}$ The table in Figure 3-24 shows that normal hearing sensitivity at 250 Hz is 25.5 dB_{SPL}, meaning that the listener has slightly better than normal hearing at this frequency. As a final example, imagine that this listener barely detects a 500 Hz tone at 30 dB_{SPL}. Since the table shows that normal hearing sensitivity at 500 Hz is 11.5 dB_{SPL}, the listener has a hearing loss of 30.0-11.5 = 18.5 dB. The simple point to be made about these examples is that, with an audiometer dial that is calibrated in dB_{SPL}, it is not possible to determine whether a listener has a hearing loss, or to measure the size of that loss, without doing some arithmetic involving the normative data in Figure 3-24. The dB_{HL} scale, however, provides a simple solution to this problem that avoids this arithmetic entirely. The solution involves calibrating the audiometer in such a way that, when the level dial is set to 0 dB_{HL}, sound level is set to the threshold of audibility for the average normal-hearing listener *for that signal frequency.* For example, when the level dial is set to 0 dB_{HL} at 125 Hz the level of tone will be 45 dB_{SPL} – the threshold of audibility for the average normal hearing listener at this frequency. Now if a listener barely detects the 125 Hz tone at 0 dB_{HL}, no arithmetic is needed; the listener has normal hearing at this frequency. Further, if the listener barely detects this 125 Hz tone at 40 dB_{HL}, for example, the listener must have a 40 dB loss at this frequency – and again it is not necessary to consult the data in Figure 3-24. Similarly, when the level dial is set to 0 dB_{HL} at 250 Hz the level of the tone will be 25.5 dB_{SPL}, which is the audibility threshold at 250 Hz. If this tone is barely detected at 0 dB_{HL}, the listener has normal hearing at this frequency. However, if the tone is not heard until the dial is increased to 50 dB dB_{HL}, for example, the listener has a 50 dB hearing loss at this frequency. The same system is used for all signal frequencies: in all cases, the 0 dB_{HL} reference is not a fixed number as it is for dB_{SPL} (a constant value of 20 μPa, no matter what the signal frequency is) or dB_{IL} (a constant value of 10^{-12} watts/m^2, again independent of signal frequency), but rather a family of numbers. In each case the reference for the dB_{HL} scale is *the threshold of audibility for an average, normal-hearing listener at a particular signal frequency.* What this means is that values in dB_{HL} are *a fixed distance above the audibility curve*, although they may be very different levels in dB_{SPL}. For illustration, Figure 3-25 shows the audibility curve (the filled symbols) and, above that in the unfilled symbols, a collection of values that all measure 30 dB_{HL}. Although the sound levels on the 30 dB_{HL} curve vary considerably in dB_{SPL} (i.e. measured using 20 μPa as the reference), every data point on this curve is a constant 3 factors of 10, or 30 dB, *above the audibility curve.* The value of 30 dB in this figure is just an example. All values in dB_{HL} and dB_{SPL} are interpreted in the same way: 50 dB_{SPL} means that the signal being measured is 100,000 times (i.e., 5 factors of 10) more intense than the fixed reference of 20 μPa, independent of frequency; 50 dB_{HL}, on the other hand, means that the signal being measured is 100,000 times (again, 5 factors of 10) more intense than a tone that is barely audible to a normal-hearing listener at that signal frequency. Similarly, 20 dB_{SPL} means that the signal is 20 dB (2 factors of 10) more intense than the fixed reference of 20 μPa, while 20 dB_{HL} means that the signal is 20 dB (again, 2 factors of 10) above the audibility curve.

[6] A clinical audiometer is an instrument with, among other things, one dial (for each ear) that controls pure-tone frequency and another dial that controls the intensity of the tone. The listener is asked to raise a hand when the tone is barely audible.

Summary

The decibel is a powerful scale for representing signal amplitude. The scale has two important properties: (1) similar to the Mach scale, it represents signal level not in absolute terms but as a measured level divided by a reference level; and (2) like the Richter scale, the dB scale is logarithmic rather than linear, meaning that it is based on equal multiplicative distances rather than equal additive distances. While the decibel is defined in terms of intensity ratios, for practical reasons, measures of sound pressure are far more common than measures of sound intensity. Consequently, a version of the decibel formula was derived that makes use of pressure ratios rather than intensity ratios. The derivation was based on the fact that intensity is proportional to pressure squared. The two versions of the decibel formula ($dB_{IL} = 10 \log_{10} I_m/I_r$ and $dB_{SPL} = 20 \log_{10} E_m/E_r$) are fully equivalent, meaning that if a sound measures 60 dB_{IL} that same sound will measure 60 dB_{SPL}. Unlike the Mach scale, which always uses the speed of sound as a reference, any number of references can be used with the decibel scale. The standard reference for the dB_{IL} scale is 10^{-12} w/m^2 and the standard reference for the dB_{SPL} scale is 20 μPa. However, any level can be used as a reference as long as it is specified. The dB_{HL} scale, widely used in audiological assessment, was developed specifically for measuring sensitivity to pure tones of difference frequencies. The reference that is used for the dB_{HL} scale is the threshold of audibility at a particular signal frequency for the average, normal-hearing listener. Sound levels in dB_{SPL} and dB_{HL} are interpreted quite differently. For example, a pure tone measuring 40 dB_{SPL} is 4 factors of 10 (i.e., 40 dB) greater than the fixed SPL reference of 20 μPa, while a pure tone measuring 40 dB_{HL} is 4 factors of 10 (again, 40 dB) greater than a tone of that same frequency that is barely audible to an average, normal-hearing listener.



Frequency	Threshold
125	45.0
250	25.5
500	11.5
750	8.0
1000	7.0
1500	6.5
2000	9.0
3000	10.0
4000	9.5
6000	15.5
8000	13.0

Figure 3-24. The threshold of audibility for the average, normal-hearing listener for pure tones varying between 125 and 8000 Hz. The audibility threshold is the sound level in dB_{SPL} that is required for a listener to barely detect a tone. Values on this curve are shown in the table to the right. The most important point to note about this graph is that the curve is not flat, meaning that the ear is more sensitive at some frequencies than others. In particular, the ear is more sensitive in a range of mid-frequencies between about 1000 and 4000 Hz than it is at lower and higher frequencies. The complex shape of this curve provides the underlying motivation for the dB_{HL} scale. See text for details.

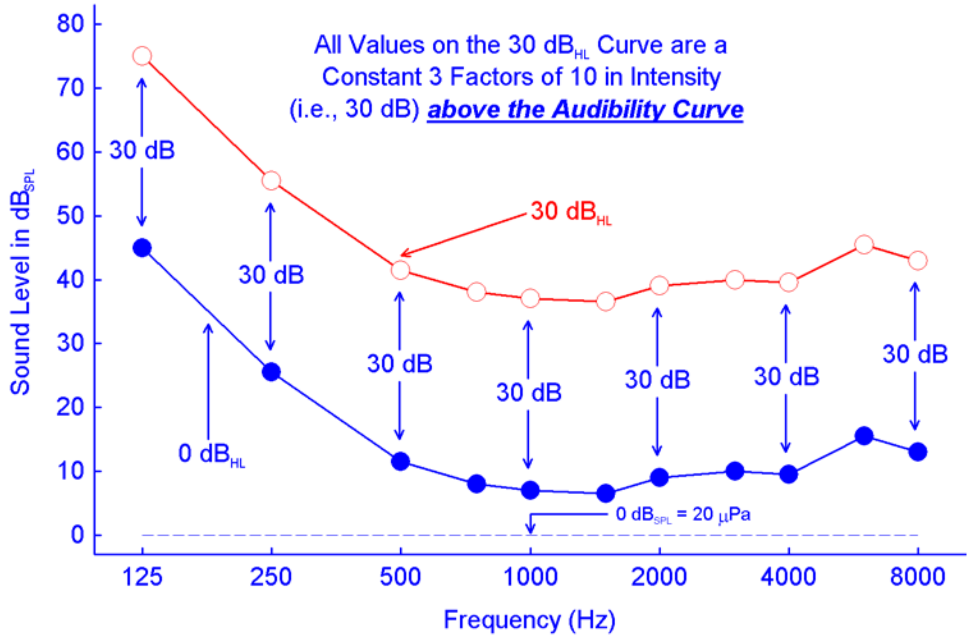


Figure 3-25. The lower function is the audibility curve – the sound level in dB_{SPL} that is required for an average normal hearing listener to barely detect pure tones of different frequencies. The upper function shows sound levels for a set of tones that all measure 30 dB_{HL}. These tones vary quite a bit in dB_{SPL} (i.e., relative to the constant value of 20 μPa) but in all cases the tones are a constant 3 factors of 10 in intensity (i.e., 30 dB) ***above the audibility curve***.

Box 3-3

THE EQUIVALENCE OF THE INTENSITY AND PRESSURE VERSIONS OF THE DECIBEL FORMULA

One fact about the two versions of the dB formula that is not always well understood is that *the dB_{IL} and dB_{SPL} formulas are fully equivalent*. By "fully equivalent" we mean the following: suppose that a sound intensity meter is used to measure the level of some sound, and we find that this sound is 1,000 times more intense than the standard intensity reference of 10^{-12} w/m^2. The sound would then measure 30 dB_{IL} (10 $\log_{10}$ 1,000 = 10 (3) = 30 dB_{IL}). Now suppose that we put the sound intensity meter away and use a sound pressure meter to measure the same sound. You might think that the sound would measure 60 dB_{SPL} since now we are multiplying by 20 instead of 10, but the trick is that *the ratio is no longer 1,000*. Recall that intensity is proportional to pressure squared, which means that pressure is proportional to the square root of intensity. This means that if the intensity ratio is 1,000, the pressure ratio must be the square root of 1,000, or 31.6. So, the formula now becomes 20 log 31.6 = 20 (1.5) = 30 dB_{SPL}, which is exactly what we obtained originally. It will always work out this way: if a sound measures 50 dB_{IL}, that same sound will measure 50 dB_{SPL}.

Table 3-2 might help to make this more clear. The first column shows an intensity ratio, the second column shows the corresponding pressure ratio (this is always the square root of the intensity ratio), the third column shows the dB_{IL} value (10 log of the intensity ratio), and the fourth column shows dB_{SPL} value (20 log of the pressure ratio). As you can see, they are always the same.

Table 3-2. Intensity ratios, equivalent pressure ratios, dB_{IL} values and dB_{SPL} values showing the equivalence of the intensity and pressure versions of the dB formula.

Intensity Ratio	Pressure Ratio	dB_{IL} (10 $\log_{10} I_m/I_r$)	dB_{SPL} (20 $\log_{10} E_m/E_r$)
10	3.16	10.00	10.00
20	4.47	13.01	13.01
40	6.32	16.02	16.02
50	7.07	16.99	16.99
60	7.75	17.78	17.78
70	8.37	18.45	18.45
80	8.94	19.03	19.03
90	9.49	19.54	19.54
100	10.00	20.00	20.00
200	14.14	23.01	23.01
300	17.32	24.77	24.77
400	20.00	26.02	26.02
500	22.36	26.99	26.99
1000	31.62	30.00	30.00

Study Questions: Physical Acoustics

1. Explain the basic processes that are involved in the propagation of a sound wave.

2. Draw time- and frequency-domain representations of simple periodic, complex periodic, complex aperiodic, and transient sounds.

3. Draw time- and frequency-domain representations of two complex periodic sounds with different fundamental frequencies.

4. Draw time-domain representations of two simple periodic sounds with the same frequency and phase, but different amplitudes.

5. Draw time-domain representations of two simple periodic sounds with the same frequency and different amplitudes but different phases.

6. Draw amplitude spectra of two sounds with the same fundamental frequencies but different spectrum envelopes.

7. Draw amplitude spectra of two sounds with different fundamental frequencies but similar spectrum envelopes.

8. Calculate signal frequencies for sinusoids with the following values:

    a. period = 0.34 s
    b. period = 2 s
    c. period = 10 ms
    d. period = 2 ms
    e. wavelength = 20 cm
    f. wavelength = 100 cm

Answers:

    a. f = 1/0.34 = 2.94 Hz
    b. f = 1/2 = 0.5 Hz
    c. f = 1/0.01 = 100 Hz
    d. f = 1/.002 = 500 Hz
    e. f = c/WL (speed of sound/wavelength) = 35000/20 = 1750 Hz
    f. f = c/WL (speed of sound/wavelength) = 35000/100 = 350 Hz

9. Calculate the three lowest resonant frequencies of the following uniform tubes that are closed at one end and open at the other end:

    a. 10 cm
    b. 30 cm
    c. 40 cm

Answers:

a. wavelength of lowest resonance = 40 cm (10 x 4)
f = 35000/40 = 875
R1 = 875 (R1 = frequency of resonance number 1)
R2 = 2625
R3 = 4375

b. wavelength of lowest resonance = 120 cm (30 x 4)
f = 35000/120 = 291.7

R1 = 291.7
R2 = 875.0
R3 = 1458.3

c. wavelength of lowest resonance = 160 cm (40 x 4)
f = 35000/160 = 218.75
R1 = 218.75
R2 = 656.25
R3 = 1093.75

10. Show what the frequency-response curves look like for the tubes in the problem above.

11. A complex periodic signal has a fundamental period of 4 msec. What is the fundamental frequency of the signal? At what frequencies would we expect to find energy?

12. How are the terms *octave* and *harmonic* different?

13. Give examples of the following kinds of graphs, being sure to label both axes:

a. amplitude spectrum
b. phase spectrum
c. frequency-response curve
d. time-domain representation

14. Give a brief explanation of the basic idea behind Fourier analysis. What is the input to Fourier analysis and what kind of output(s) does it produce?

15. Draw and label frequency-response curves for low-pass, high-pass, and band-pass filters.

16. What parameters control the frequency of vibration of a spring and mass system?

17. Draw the time domain representation of one cycle of a sinusoid as variations in instantaneous air pressure over time and one cycle of that same sinusoid as variations in instantaneous velocity over time.

18. How, if at all, are the terms *resonant frequency* and *harmonic* different?

19. How, if at all, are the terms *resonant frequency* and *formant* different?

20. A harmonic is a peak in: (a) a frequency response curve, (b) an amplitude spectrum, or (c) either a frequency response curve or an amplitude spectrum.

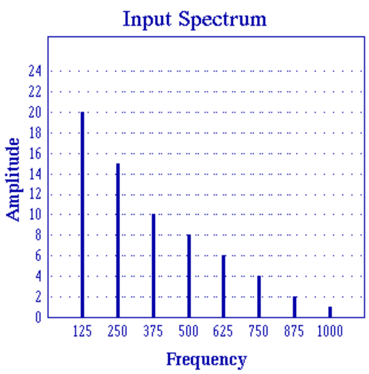
21. A resonance is a peak in: (a) a frequency response curve, (b) an amplitude spectrum, or (c) either a frequency response curve or an amplitude spectrum.

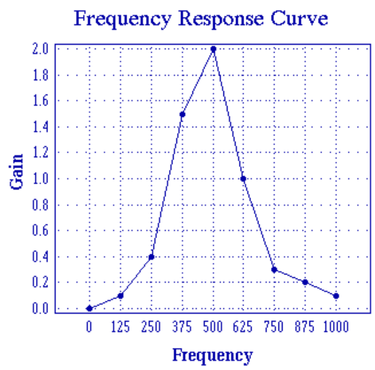
22. A formant is a peak in: (a) a frequency response curve, (b) an amplitude spectrum, or (c) either a frequency response curve or an amplitude spectrum.

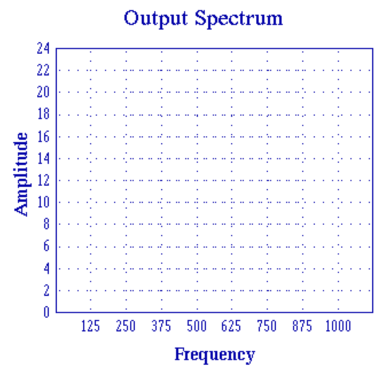
23. A frequency response curve describes a __________________________.

24. An amplitude spectrum describes a __________________________.

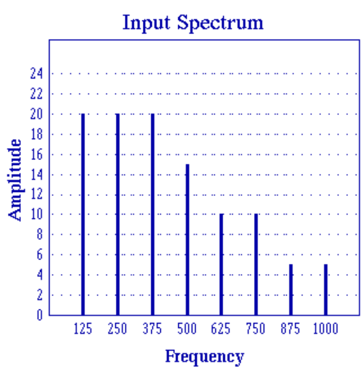
Frequency Response Problems

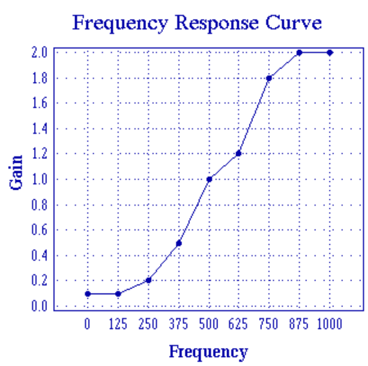


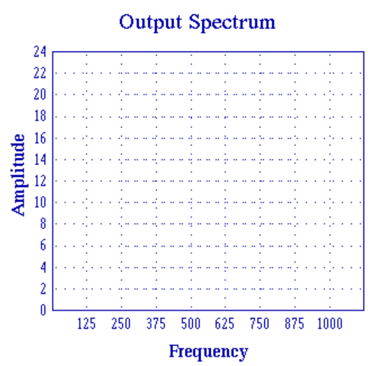




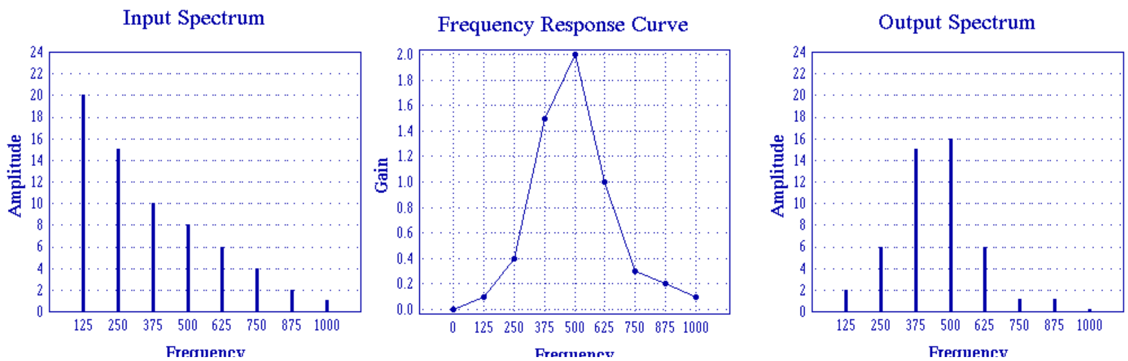
Assume that a signal with the amplitude spectrum show at the left is modified by a filter with the frequency response curve show in the middle. Show what the output spectrum would look like.

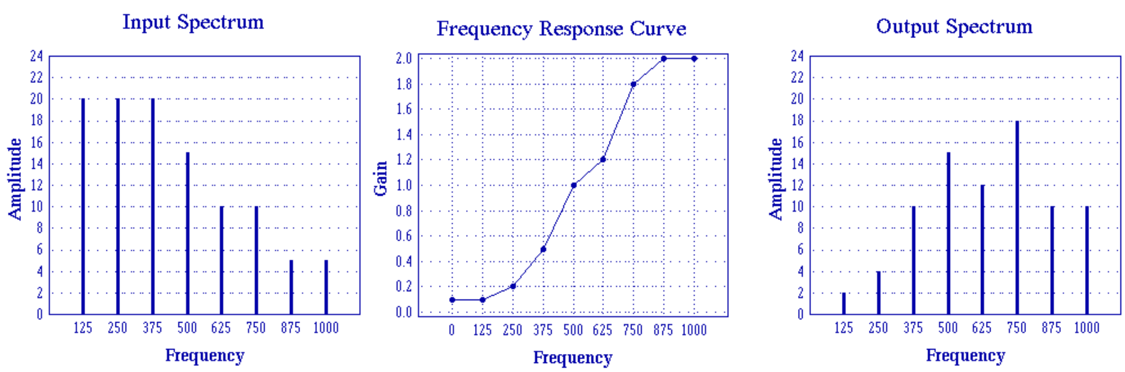






<u>Answers to Frequency Response Problems</u>





Decibel Study Questions

1. What reference is used for the dB_{IL} scale?

2. What reference is used for the dB_{SPL} scale?

3. What reference is used for the dB_{HL} scale?

4. What reference is used for the dB_{SL} scale?

5. A listener barely detects a 125 Hz pure tone at 55 dB_{SPL}. Does this listener have a hearing loss at 125 Hz, and if so, what is the size of the hearing loss?

6. A listener barely detects a 1,000 Hz pure tone at 55 dB_{SPL}. Does this listener have a hearing loss at 1,000 Hz, and if so, what is the size of the hearing loss?

7. A listener barely detects a 125 Hz pure tone at 55 dB_{HL}. Does this listener have a hearing loss at 125 Hz, and if so, what is the size of the hearing loss?

8. A listener barely detects a 1,000 Hz pure tone at 55 dB_{HL}. Does this listener have a hearing loss at 1,000 Hz, and if so, what is the size of the hearing loss?

9. 60 dB_{SPL} at 1,000 Hz means ____________________ more intense than ____________________.

10. 60 dB_{IL} at 1,000 Hz means ____________________ more intense than ____________________.

11. 60 dB_{HL} at 1,000 Hz means ____________________ more intense than ____________________.

12. The reference that is used for the dB_{SPL} scale is:

    a. a number
    b. a sentence

13. If the answer to the question above is a number, give the number; if it's a sentence, give the sentence.

14. The reference that is used for the dB_{HL} scale is:

    a. a number
    b. a sentence

15. If the answer to the question above is a number, give the number; if it's a sentence, give the sentence.

16. A specific individual has a 70 dB hearing loss in the left ear at 1,000 Hz. A 90 dB_{HL}, 1,000 Hz tone that is presented to this listener's left ear would measure ______ dB_{SL}.

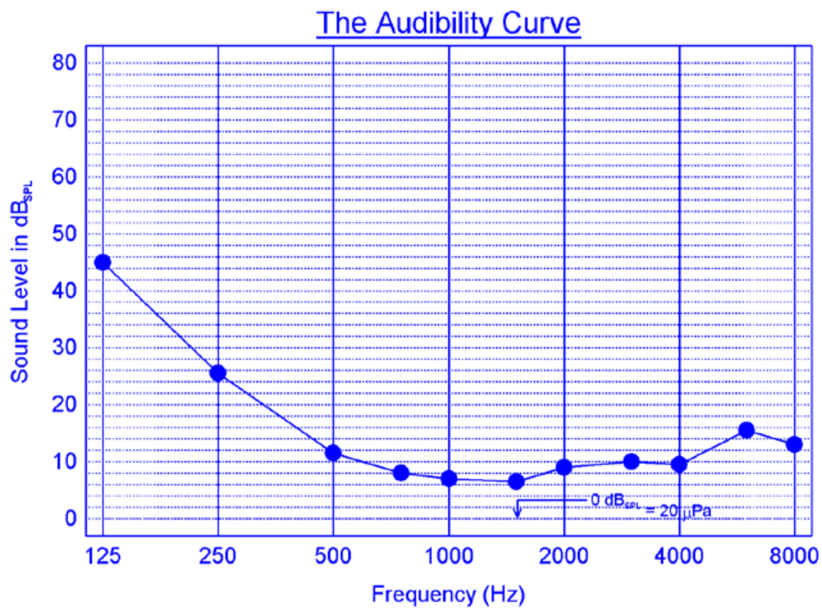
17. A sound measures 42 dB_{IL}. On the dB_{SPL} scale, that same sound will measure:

    a. 84 dB_{SPL} because with the dB_{SPL} formula we are now are multiplying the ratio by 20 instead of 10.
    b. 42 dB_{SPL} because the two versions of the formula are equivalent

18. A sound measures 60 dB_{IL}. (a) The measured intensity (I_M) must therefore be _________ times greater than the reference intensity (I_R). (b) What would the pressure ratio (E_M/E_R) be for this same sound? (c) Do the arithmetic to show what this sound would measure in dB_{SPL}.

19. A sound measures 40 dB_{IL}. (a) The measured intensity (I_M) must therefore be _________ times greater than the reference intensity (I_R). (b) What would the pressure ratio (E_M/E_R) be for this same sound? (c) Do the arithmetic to show what this sound would measure in dB_{SPL}.

20. On the graph below, put a mark at: (a) 3,000 Hz, 20 dB_{SPL}, and (b) 3,000 Hz, 20 dB_{HL} (the grid lines on the y axis are spaced at 2 dB intervals).



Frequency in Hz	Threshold in dB SPL
125	45.0
250	25.5
500	11.5
750	8.0
1000	7.0
1500	6.5
2000	9.0
3000	10.0
4000	9.5
6000	15.5
8000	13.0

Answers to Decibel Study Questions

1. 10^{-12} watts/m^2

2. 20 μPa (or, equivalently, 0.0002 dynes/cm^2)

3. The threshold of audibility for an average, normal-hearing listener at a particular signal frequency.

4. 3. The threshold of audibility for a particular listener at a particular signal frequency.

5. Consulting the attached figure and table showing the audibility curve for average, normal-hearing listeners, we find that the threshold of audibility at 125 Hz is 45 dB_{SPL}. A listener who barely detected a 125 Hz tone at 55 dB_{SPL} would therefore have hearing loss of 55-45=10 dB; that is, the hearing sensitivity of this listener would be 10 dB worse than normal.

6. Consulting the attached figure and table showing the audibility curve for average, normal-hearing listeners, we find that the threshold of audibility at 1,000 Hz is 7 dB_{SPL}. A listener who barely detected a 1,000 Hz tone at 55 dB_{SPL} would therefore have a hearing loss of 55-7=48 dB; that is, the hearing sensitivity of this listener would be 48 dB worse than normal.

7. The reference for dB_{HL} *is* the audibility threshold, so this listener would have a 55 dB hearing loss at 125 Hz. There is no need to consult the table.

8. The reference for dB_{HL} *is* the audibility threshold, so this listener would have a 55 dB hearing loss at 1,000 Hz. There is no need to consult the table.

9. 6 factors of 10 (i.e., 1,000,000 times) more intense than 20 μPa)

10. 6 factors of 10 (i.e., 1,000,000 times) more intense than 10^{-12} watts/m^2

11. 6 factors of 10 (i.e., 1,000,000 times) more intense than a 1,000 Hz tone that is barely audible to an average, normal-hearing listener.

12. a number

13. 20 μPa

14. a sentence

15. The threshold of audibility for an average, normal-hearing listener at a particular signal frequency.

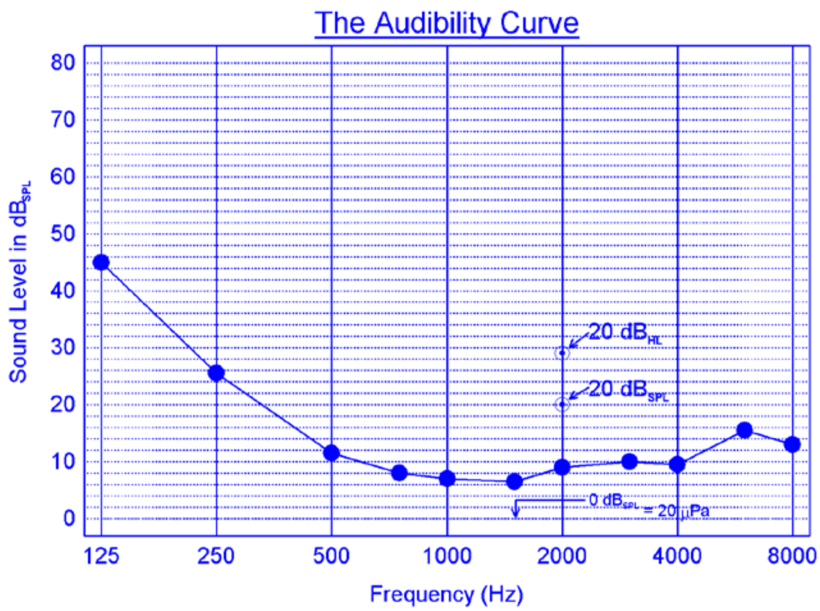
16. 20 dB_{SL}. The reference for the dB_{SL} (SL=sensation level) is the threshold of audibility for a specific listener. So, what we want to know here very simply is where this 90 dB_{HL} tone is in relation to *this particular listener's* threshold. This listener has a 70 dB hearing loss at this frequency, so the 90 dBHL tone, which would be 90 dB above a normal-hearing listener's threshold, is only 20 dB above this particular listener's threshold.

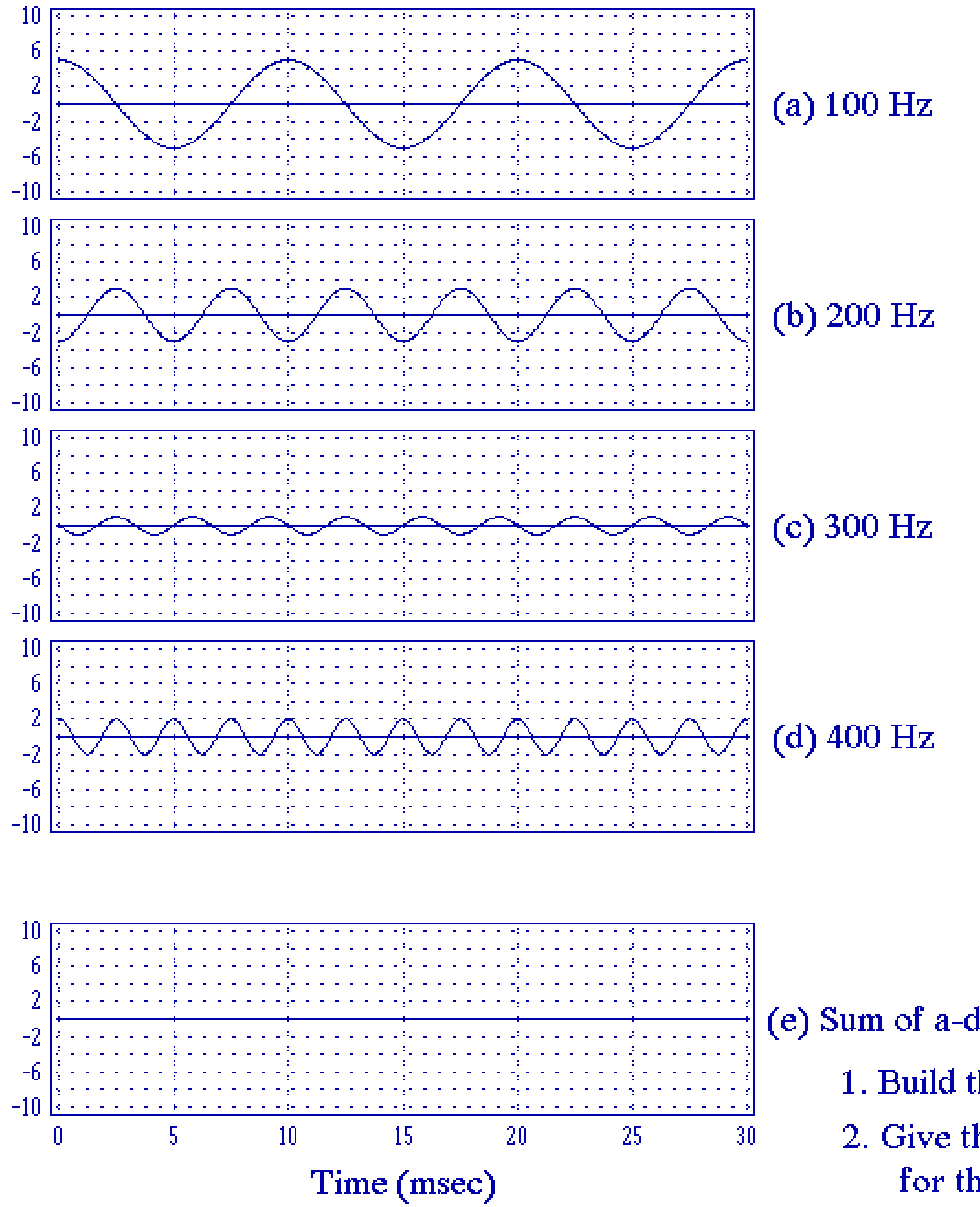
17. 42 dB_{SPL}: The pressure version of the formula was derived from the intensity version through algebraic manipulations, so they have to be equivalent to one another. The next problem was designed to illustrate how this can be the case.

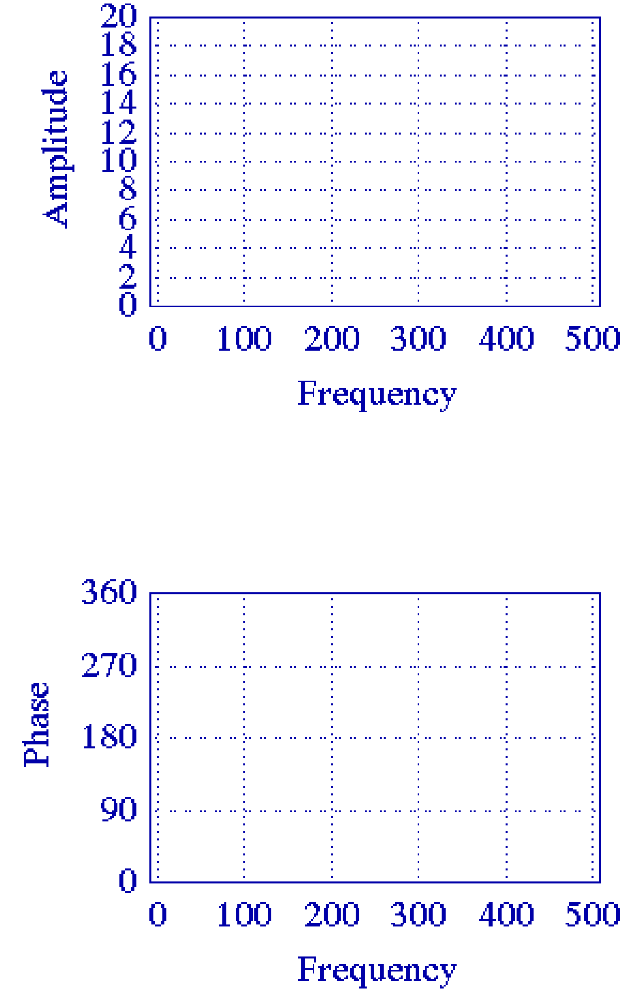
18. (a) 1,000,000 times (6 factors of 10) more intense than I_R. (b) If the intensity ratio is 1,000,000, the pressure ratio has to be the square root of 1,000,000, which is 1,000. (c) dB_{SPL} = 20 log 1,000 = 20 $\cdot$ 3 = 60 dB_{SPL}. This is exactly what we got for the same sound measured in dB_{IL}. It will always be the same. <u>If a sound measures 60 dB_{IL}, *that same sound* will measure 60 dB_{SPL}.</u>

19. (a) 10,000 times (4 factors of 10) more intense than I_R. (b) If the intensity ratio is 10,000, the pressure ratio has to be the square root of 10,000, which is 100. (c) $dB_{SPL} = 20 \log 100 = 20 \cdot 2 = 40$ dB_{SPL}. This is exactly what we got for the same sound measured in dB_{IL}. It will always be the same. <u>If a sound measures 40 dB_{IL}, *that same sound* will measure 40 dB_{SPL}</u>.

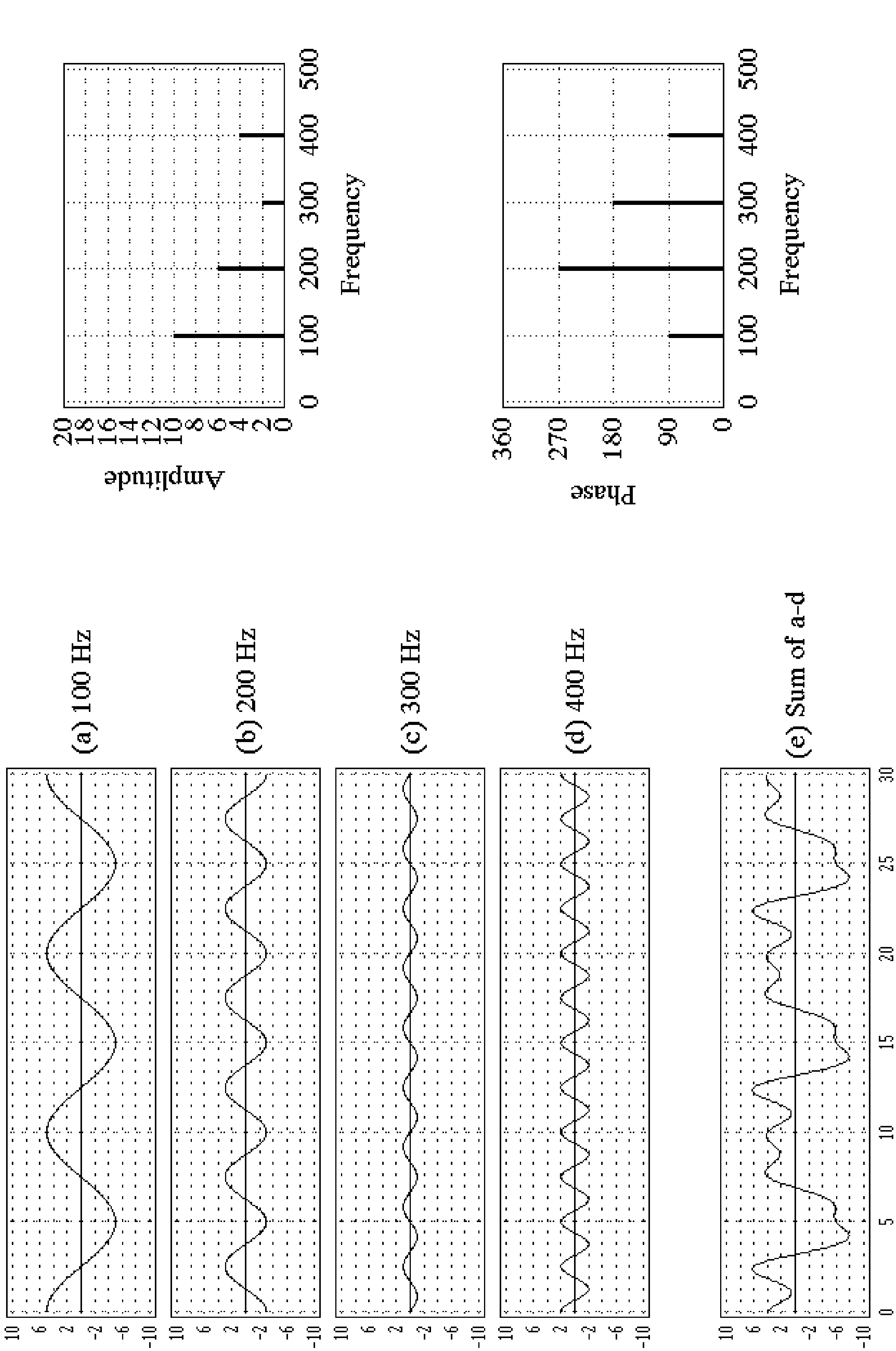
20. See below. The lower of the two marks is 20 dB (2 factors of 10) above the constant reference line of 20 μPa. The higher of the two marks is 20 dB (also 2 factors of 10) above the *curvey* line, which is the threshold of audibility for the average normal-hearing listener.







1. Build the complex wave (e) from components (a)-(d).
2. Give the amplitude spectrum and the phase spectrum for the complex periodic signal (e).



Answer to Fourier Analysis Problem