Chapter

1.1

The Physical Nature of Sound

Floyd E. Toole

E. A. G. Shaw, G. A. Daigle, M. R. Stinson

1.1.1 Introduction

Sound is a physical disturbance in the medium through which it is propagated. Although the most common medium is air, sound can travel in any solid, liquid, or gas. In air, sound consists of localized variations in pressure above and below normal atmospheric pressure (*compressions* and *rarefactions*).

Air pressure rises and falls routinely, as environmental weather systems come and go, or with changes in altitude. These fluctuation cycles are very slow, and no perceptible sound results, although it is sometimes evident that the ears are responding in a different way to these *infrasonic* events. At fluctuation frequencies in the range from about 20 cycles per second up to about 20,000 cycles per second the physical phenomenon of sound can be perceived as having pitch or tonal character. This generally is regarded as the *audible* or *audio-frequency range*, and it is the frequencies in this range that are the concern of this chapter. Frequencies above 20,000 cycles per second are classified as *ultrasonic*.

1.1.2 Sound Waves

The essence of sound waves is illustrated in Figure 1.1.1, which shows a tube with a piston in one end. Initially, the air within and outside the tube is all at the prevailing atmospheric pressure. When the piston moves quickly inward, it compresses the air in contact with its surface. This energetic compression is rapidly passed on to the adjoining layer of air, and so on, repeatedly. As it delivers its energy to its neighbor, each layer of air returns to its original uncompressed state. A longitudinal sound pulse is moving outward through the air in the tube, causing only a passing disturbance on the way. It is a pulse because there is only an isolated action, and it is longitudinal because the air movement occurs along the axis of sound propagation. The rate at which the pulse propagates is the speed of sound. The pressure rise in the compressed air is proportional to the velocity with which the piston moves, and the perceived loudness of the resulting sound pulse

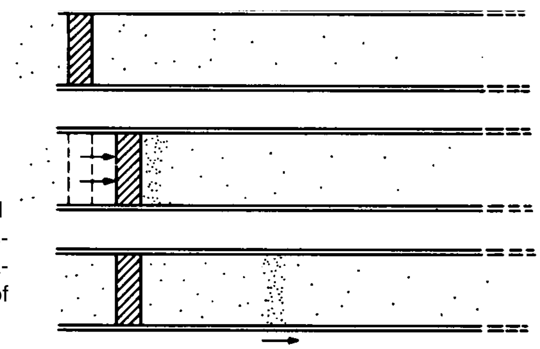


Figure 1.1.1 Generation of a longitudinal sound wave by the rapid movement of a piston in the end of a tube, showing the propagation of the wave pulse at the speed of sound down the length of the tube.

is related to the incremental amplitude of the pressure wave above the ambient atmospheric pressure.

Percussive or impulsive sounds such as these are common, but most sounds do not cease after a single impulsive event. Sound waves that are repetitive at a regular rate are called *periodic.* Many musical sounds are periodic, and they embrace a very wide range of repetitive patterns. The simplest of periodic sounds is a pure tone, similar to the sound of a tuning fork or a whistle. An example is presented when the end of the tube is driven by a loudspeaker reproducing a recording of such a sound (Figure 1.1.2). The pattern of displacement versus time for the loudspeaker diaphragm, shown in Figure 1.1.2*b*, is called a *sine wave* or *sinusoid*.

If the first diaphragm movement is inward, the first event in the tube is a pressure compression, as seen previously. When the diaphragm changes direction, the adjacent layer of air undergoes a *pressure rarefaction*. These cyclic compressions and rarefactions are repeated, so that the sound wave propagating down the tube has a regularly repeated, periodic form. If the air pressure at all points along the tube were measured at a specific instant, the result would be the graph of air pressure versus distance shown in Figure 1.1.2*c*. This reveals a smoothly sinusoidal waveform with a repetition distance along the tube symbolized by λ, the *wavelength* of the periodic sound wave.

If a pressure-measuring device were placed at some point in the tube to record the instantaneous changes in pressure at that point as a function of time, the result would be as shown in Figure 1.1.2*d*. Clearly, the curve has the same shape as the previous one except that the horizontal axis is time instead of distance. The periodic nature of the waveform is here defined by the time period T, known simply as the *period* of the sound wave. The inverse of the period, $1/T$, is the *frequency* of the sound wave, describing the number of repetition cycles per second passing a fixed point in space. An ear placed in the path of a sound wave corresponding to the musical tone middle C would be exposed to a frequency of 261.6 cycles per second or, using standard scientific terminology, a frequency of 261.6 hertz (Hz). The perceived loudness of the tone would depend on the magnitude of the pressure deviations above and below the ambient air pressure.

The parameters discussed so far are all related by the *speed of sound.* Given the speed of sound and the duration of one period, the wavelength can be calculated as

$$\lambda = cT \tag{1.1.1}$$

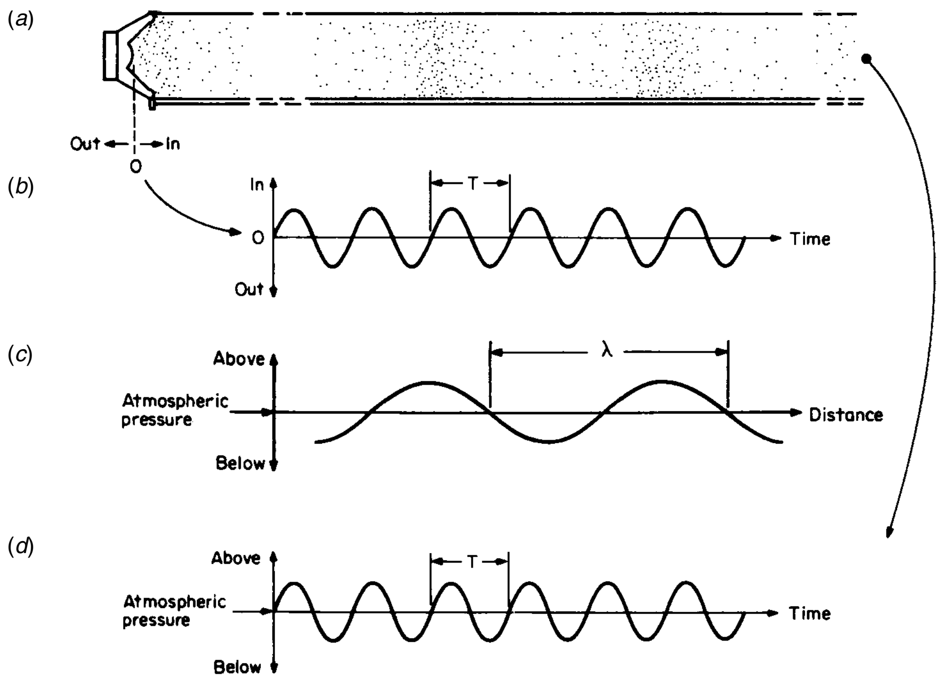


Figure 1.1.2 Characteristics of sound waves: (*a*) A periodic sound wave, a sinusoid in this example, is generated by a loudspeaker placed at the end of a tube. (*b*) Waveform showing the movement of the loudspeaker diaphragm as a function of time: displacement versus time. (*c*) Waveform showing the instantaneous distribution of pressure along a section of the tube: pressure versus distance. (*d*) Waveform showing the pressure variation as a function of time at some point along the tube: pressure versus time.

where:
λ = wavelength
c = speed of sound
T = period

By knowing that the frequency $f = 1/T$, the following useful equation and its variations can be derived:

$$\lambda = \frac{c}{f} \qquad f = \frac{c}{\lambda} \qquad c = f\lambda \tag{1.1.2}$$

The speed of sound in air at a room temperature of 22°C (72°F) is 345 m/s (1131 ft/s). At any other ambient temperature, the speed of sound in air is given by the following approximate relationships [1, 2]:

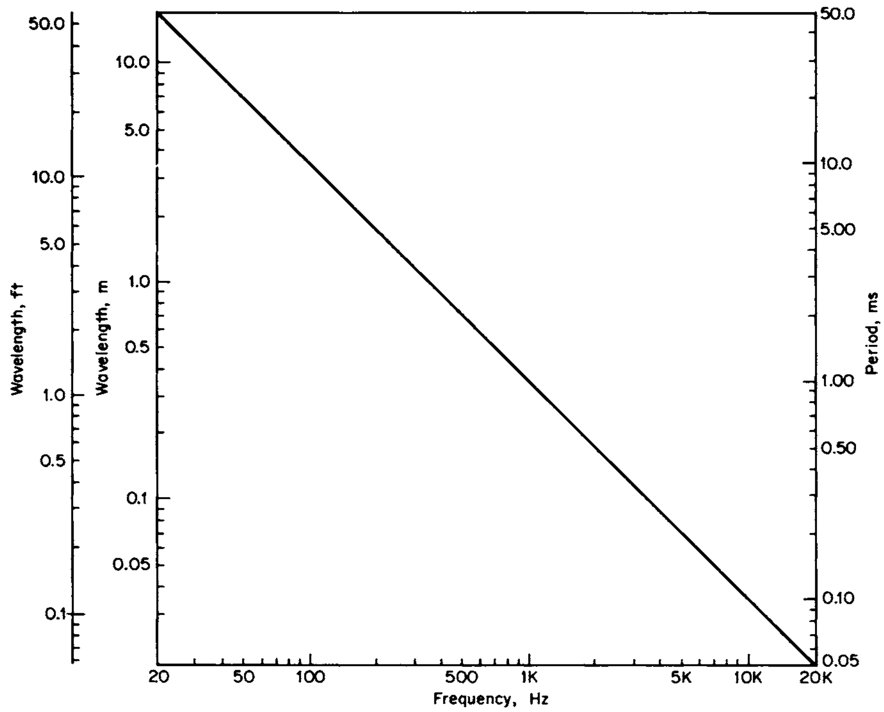


Figure 1.1.3 Relationships between wavelength, period, and frequency for sound waves in air.

$$c(m/s) = 331.29 + 0.607t(°C) \tag{1.1.3}$$

or

$$c(m/s) = 1051.5 + 1.106t(°F) \tag{1.1.4}$$

where t = ambient temperature.

The relationships between the frequency of a sound wave and its wavelength are essential to understanding many of the fundamental properties of sound and hearing. The graph of Figure 1.1.3 is a useful quick reference illustrating the large ranges of distance and time embraced by audible sounds. For example, the tone middle C with a frequency of 261.6 Hz has a wavelength of 1.3 m (4.3 ft) in air at 20°C. In contrast, an organ pedal note at Cl, 32.7 Hz, has a wavelength of 10.5 m (34.5 ft), and the third-harmonic overtone of C8, at 12,558 Hz, has a wavelength of 27.5 mm (1.1 in). The corresponding periods are, respectively, 3.8 ms, 30.6 ms, and 0.08 ms. The contrasts in these dimensions are remarkable, and they result in some interesting and troublesome effects in the realms of perception and audio engineering. For the discussions that follow it is often more helpful to think in terms of wavelengths rather than in frequencies.

1.1.2a Complex Sounds

The simple sine waves used for illustration reveal their periodicity very clearly. Normal sounds, however, are much more complex, being combinations of several such pure tones of different frequencies and perhaps additional transient sound components that punctuate the more sustained elements. For example, speech is a mixture of approximately periodic vowel sounds and staccato consonant sounds. Complex sounds can also be periodic; the repeated wave pattern is just more intricate, as is shown in Figure 1.l.4*a*. The period identified as T_1 applies to the *fundamental frequency* of the sound wave, the component that normally is related to the characteristic pitch of the sound. Higher-frequency components of the complex wave are also periodic, but because they are typically lower in amplitude, that aspect tends to be disguised in the summation of several such components of different frequency. If, however, the sound wave were analyzed, or broken down into its constituent parts, a different picture emerges: Figure 1.l.4*b*, *c*, and *d.* In this example, the analysis shows that the components are all *harmonics,* or whole-number multiples, of the fundamental frequency; the higher-frequency components all have multiples of entire cycles within the period of the fundamental.

To generalize, it can be stated that all *complex periodic waveforms* are combinations of several harmonically related sine waves. The shape of a complex waveform depends upon the relative amplitudes of the various harmonics and the position in time of each individual component with respect to the others. If one of the harmonic components in Figure 1.1.4 is shifted slightly in time, the shape of the waveform is changed, although the frequency composition remains the same (Figure 1.1.5). Obviously a record of the time locations of the various harmonic components is required to completely describe the complex waveform. This information is noted as the *phase* of the individual components.

1.1.2b Phase

Phase is a notation in which the time of one period of a sine wave is divided into 360°. It is a relative quantity, and although it can be defined with respect to any reference point in a cycle, it is convenient to start (0°) with the upward, or positive-going, zero crossing and to end (360°) at precisely the same point at the beginning of the next cycle (Figure 1.1.6). *Phase shift* expresses in degrees the fraction of a period or wavelength by which a single-frequency component is shifted in the time domain. For example, a phase shift of 90° corresponds to a shift of one-fourth period. For different frequencies this translates into different time shifts. Looking at it from the other point of view, if a complex waveform is time-delayed, the various harmonic components will experience different phase shifts, depending on their frequencies.

A special case of phase shift is a *polarity reversal*, an inversion of the waveform, where all frequency components undergo a 180° phase shift. This occurs when, for example, the connections to a loudspeaker are reversed.

1.1.2c Spectra

Translating time-domain information into the frequency domain yields an *amplitude-frequency spectrum* or, as it is commonly called, simply a *spectrum*. Figure 1.1.7*a* shows the spectrum of the waveform in Figures 1.1.4 and 1.1.5, in which the height of each line represents the amplitude of that particular component and the position of the line along the frequency axis identifies its frequency. This kind of display is a *line spectrum* because there are sound components at only

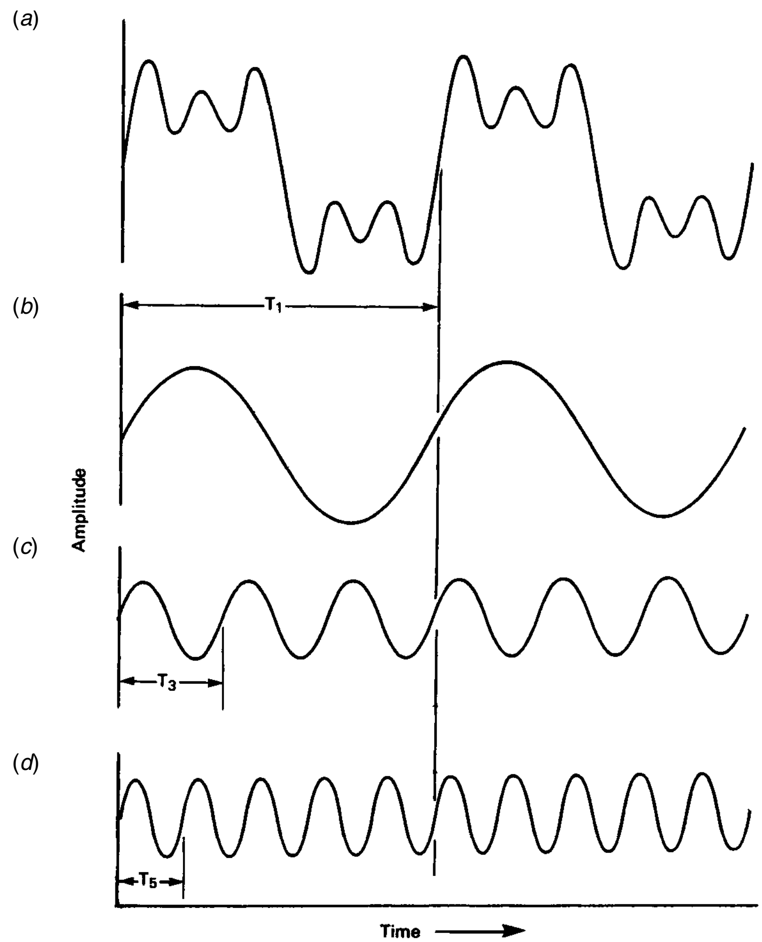


Figure 1.1.4 A complex waveform constructed from the sum of three harmonically related sinusoidal components, all of which start at the origin of the time scale with a positive-going zero crossing. Extending the series of odd-harmonic components to include those above the fifth would result in the complex waveform progressively assuming the form of a square wave. (*a*) Complex waveform, the sum of *b*, *c*, and *d*. (*b*) Fundamental frequency. (*c*) Third harmonic. (*d*) Fifth harmonic.

certain specific frequencies. The phase information is shown in Figure 1.1.7*b*, where the difference between the two waveforms is revealed in the different *phase-frequency spectra*.

The equivalence of the information presented in the two domains—the waveform in the time domain and the amplitude- and phase-frequency spectra in the frequency domain—is a matter of considerable importance. The proofs have been thoroughly worked out by the French mathematician Fourier, and the well-known relationships bear his name. The breaking down of waveforms into their constituent sinusoidal parts is known as *Fourier analysis*. The construction of complex

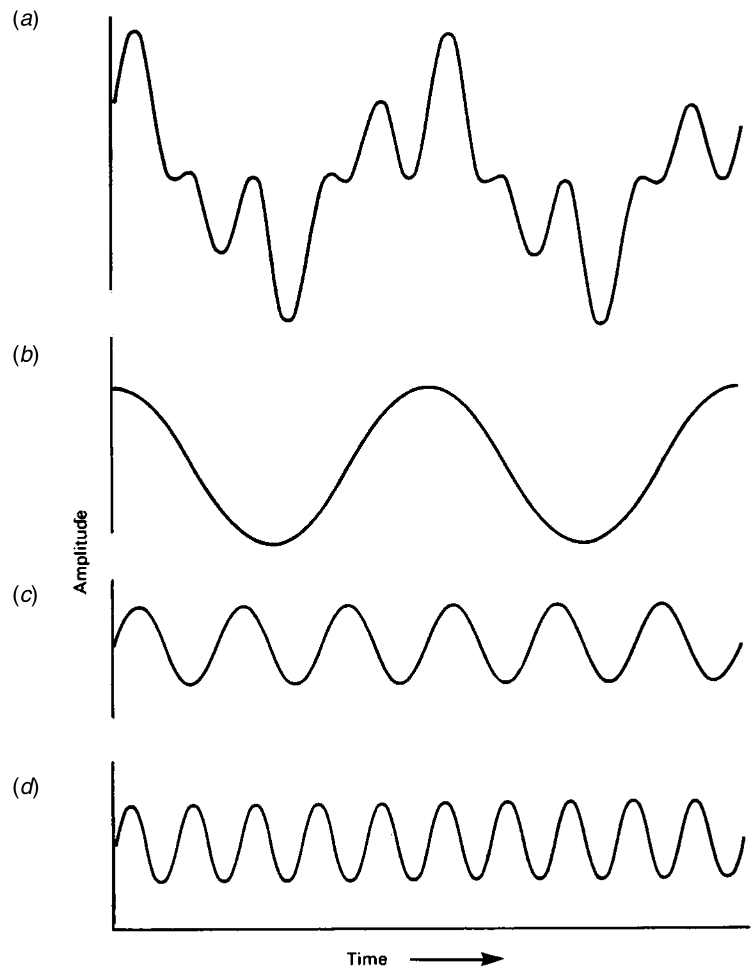


Figure 1.1.5 A complex waveform with the same harmonic-component amplitudes as in Figure 1.1.4, but with the starting time of the fundamental advanced by one-fourth period: a phase shift of 90°.

waveshapes from summations of sine waves is called *Fourier synthesis*. *Fourier transformations* permit the conversion of time-domain information into frequency-domain information, and vice versa. These interchangeable descriptions of waveforms form the basis for powerful methods of measurement and, at the present stage, provide a convenient means of understanding audio phenomena. In the examples that follow, the relationships between time-domain and frequency-domain descriptions of waveforms will be noted.

Figure 1.1.8 illustrates the sound waveform that emerges from the larynx, the buzzing sound that is the basis for vocalized speech sounds. This sound is modified in various ways in its passage down the vocal tract before it emerges from the mouth as speech. The waveform is a series

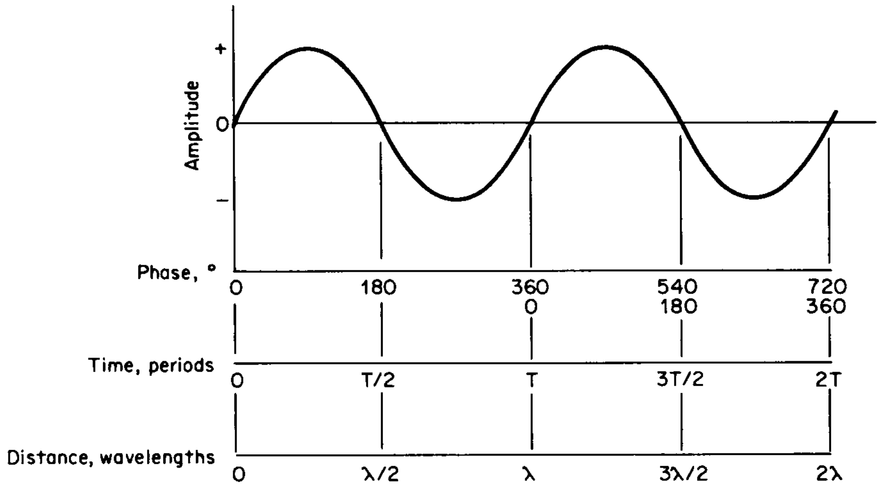


Figure 1.1.6 The relationship between the period T and wavelength λ of a sinusoidal waveform and the phase expressed in degrees. Although it is normal to consider each repetitive cycle as an independent 360°, it is sometimes necessary to sum successive cycles starting from a reference point in one of them.

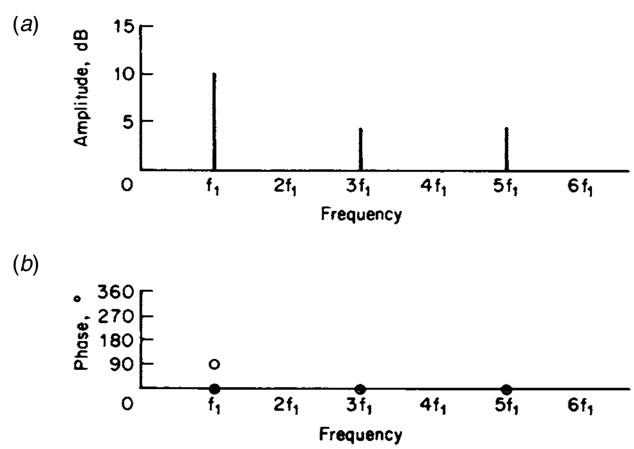


Figure 1.1.7 The amplitude-frequency spectra (*a*) and the phase-frequency spectra (*b*) of the complex waveforms shown in Figures 1.1.4 and 1.1.5. The amplitude spectra are identical for both waveforms, but the phase-frequency spectra show the 90° phase shift of the fundamental component in the waveform of Figure 1.1.5. Note that frequency is expressed as a multiple of the fundamental frequency f_1. The numerals are the harmonic numbers. Only the fundamental f_1 and the third and fifth harmonics (f_3 and f_5) are present.

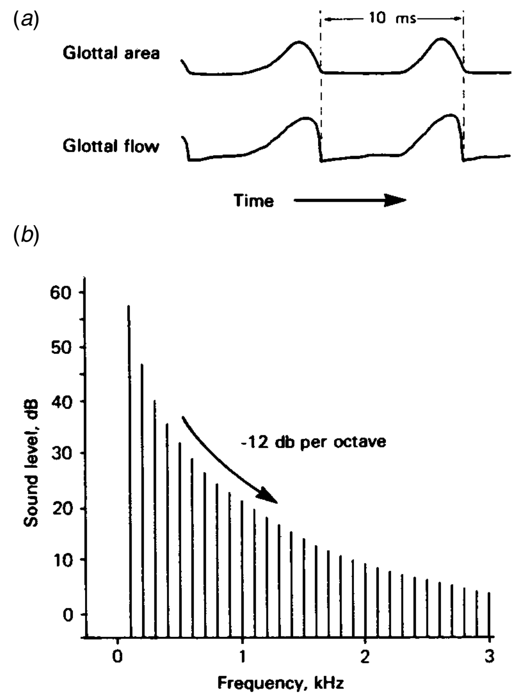


Figure 1.1.8 Characteristics of speech. (*a*) Waveforms showing the varying area between vibrating vocal cords and the corresponding airflow during vocalized speech as a function of time. (*b*) The corresponding amplitude-frequency spectrum, showing the 100-Hz fundamental frequency for this male speaker. (*From* [3]. *Used with permission.*)

of periodic pulses, corresponding to the pulses of air that are expelled, under lung pressure, from the vibrating vocal cords. The spectrum of this waveform consists of a harmonic series of components, with a fundamental frequency, for this male talker, of 100 Hz. The gently rounded contours of the waveform suggest the absence of strong high-frequency components, and the amplitude-frequency spectrum confirms it. The *spectrum envelope*, the overall shape delineating the amplitudes of the components of the line spectrum, shows a progressive decline in amplitude as a function of frequency. The amplitudes are described in *decibels,* abbreviated dB. This is the common unit for describing sound-level differences. The rate of this decline is about –12 dB per octave (an *octave* is a 2:1 ratio of frequencies).

Increasing the pitch of the voice brings the pulses closer together in time and raises the fundamental frequency. The harmonic-spectrum lines displayed in the frequency domain are then spaced farther apart but still within the overall form of the spectrum envelope, which is defined by the shape of the pulse itself. Reducing the pitch of the voice has the opposite effect, increasing the spacing between pulses and reducing the spacing between the spectral lines under the envelope. Continuing this process to the limiting condition, if it were possible to emit just a single pulse, would be equivalent to an infinitely long period, and the spacing between the spectral lines would vanish. The discontinuous, or *aperiodic,* pulse waveform therefore yields a *continuous* spectrum having the form of the spectrum envelope.

Isolated pulses of sound occur in speech as any of the variations of consonant sounds, and in music as percussive sounds and as transient events punctuating more continuous melodic lines. All these aperiodic sounds exhibit continuous spectra with shapes that are dictated by the waveforms. The leisurely undulations of a bass drum waveform contain predominantly low-frequency energy, just as the more rapid pressure changes in a snare drum waveform require the presence of higher frequencies with their more rapid rates of change. A technical waveform of considerable use in measurements consists of a very brief impulse which has the important feature of containing equal amplitudes of all frequencies within the audio-frequency bandwidth. This is moving toward a limiting condition in which an infinitely short event in the time domain is associated with an infinitely wide amplitude-frequency spectrum.

1.1.3 Dimensions of Sound

The descriptions of sound in the preceding section involved only pressure variation, and while this is the dimension that is most commonly referred to, it is not the only one. Accompanying the pressure changes are temporary movements of the air "particles" as the sound wave passes (in this context a particle is a volume of air that is large enough to contain many molecules while its dimensions are small compared with the wavelength). Other measures of the magnitude of the sound event are the displacement amplitude of the air particles away from their rest positions and the velocity amplitude of the particles during the movement cycle. In the physics of sound, the *particle displacement* and the *particle velocity* are useful concepts, but the difficulty of their measurement limits their practical application. They can, however, help in understanding other concepts.

In a normally propagating sound wave, energy is required to move the air particles; they must be pushed or pulled against the elasticity of the air, causing the incremental rises and falls in pressure. Doubling the displacement doubles the pressure change, and this requires double the force. Because the work done is the product of force times distance and both are doubled, the energy in a sound wave is therefore proportional to the square of the particle displacement amplitude or, in more practical terms, to the square of the sound pressure amplitude.

Sound energy spreads outward from the source in the three dimensions of space, in addition to those of amplitude and time. The energy of such a sound field is usually described in terms of the energy flow through an imaginary surface. The sound energy transmitted per unit of time is called *sound power*. The sound power passing through a unit area of a surface perpendicular to a specified direction is called the *sound intensity*. Because intensity is a measure of energy flow, it also is proportional to the square of the sound pressure amplitude.

The ear responds to a very wide range of sound pressure amplitudes. From the smallest sound that is audible to sounds large enough to cause discomfort, there is a ratio of approximately 1 million in sound pressure amplitude, or 1 trillion (10^{12}) in sound intensity or power. Dealing routinely with such large numbers is impractical, so a logarithmic scale is used. This is based on the *bel,* which represents a ratio of 10:1 in sound intensity or sound power (the power can be acoustical or electrical). More commonly the decibel, one-tenth of a bel, is used. A difference of 10 dB therefore corresponds to a factor-of-10 difference in sound intensity or sound power. Mathematically this can be generalized as

$$\text{Level difference} = \log \frac{P_1}{P_2} \quad \text{bels} \tag{1.1.5}$$

or

Table 1.1.1 Various Power and Amplitude Ratios and their Decibel Equivalents*

Sound or electrical power ratio	Decibels	Sound pressure, voltage, or current ratio	Decibels
1	0	1	0
2	3.0	2	6.0
3	4.8	3	9.5
4	6.0	4	12.0
5	7.0	5	14.0
6	7.8	6	15.6
7	8.5	7	16.9
8	9.0	8	18.1
9	9.5	9	19.1
10	10.0	10	20.0
100	20.0	100	40.0
1,000	30.0	1,000	60.0
10,000	40.0	10,000	80.0
100,000	50.0	100,000	100.0
1,000,000	60.0	1,000,000	120.0

* Other values can be calculated precisely by using Eqs. (1.6) and (1.7) or estimated by using this table and the following rules:

Power ratios that are multiples of 10 are converted into their decibel equivalents by multiplying the appropriate exponent by 10. For example, a power ratio of 1000 is 10^3, and this translates into $3 \times 10 = 30$ dB. Since power is proportional to the square of amplitude, the exponent of 10 must be doubled to arrive at the decibel equivalent of an amplitude ratio.

Intermediate values can be estimated by combining values in this table by means of the rule that the multiplication of power or amplitude ratios is equivalent to adding level differences in decibels. For example, increasing a sound level by 27 dB requires increasing the power by a ratio of 500 (20 dB is a ratio of 100, and 7 dB is a ratio of 5; the product of the ratios is 500). The corresponding increase in sound pressure or electrical signal amplitude is a factor of just over 20 (20 dB is a ratio of 10, and 7 dB falls between 6.0 and 9.5 and is therefore a ratio of something in excess of 2); the calculated value is 22.4. Reversing the process, if the output from a power amplifier is increased from 40 to 800 W, a ratio of 20, the sound pressure level would be expected to increase by 13 dB (a power ratio of 10 is 10 dB, a ratio of 2 is 3 dB, and the sum is 13 dB). The corresponding voltage increase measured at the output of the amplifier would be a factor of between 4 and 5 (by calculation, 4.5).

$$\text{Level difference} = 10 \log \frac{P_1}{P_2} \text{ decibels} \tag{1.1.6}$$

where P_1 and P_2 are two levels of power.

For ratios of sound pressures (analogous to voltage or current ratios in electrical systems) the squared relationship with power is accommodated by multiplying the logarithm of the ratio of pressures by 2, as follows:

$$\text{Level difference} = 10 \log \frac{P_1^2}{P_2^2} = 20 \log \frac{p_1}{p_2} \text{ dB} \tag{1.1.7}$$

where *P1* and *P2* are sound pressures.

Table 1.1.2 Typical Sound Pressure Levels and Intensities for Various Sound Sources*

Sound source	Sound pressure level, dB	Intensity, W/m^2	Listener reaction
	160		Immediate damage
Jet engine at 10 m	150	10^3	
	140		Painful feeling
	130		
SST takeoff at 500 m	120	1	Discomfort
Amplified rock music	110		
Chain saw at 1 m	100		fff
Power mower at 1.5 m	90	10^{-3}	
75-piece orchestra at 7 m	80		f
City traffic at 15 m	70		
Normal speech at 1 m	60	10^{-6}	p
Suburban residence	50		
Library	40		ppp
Empty auditorium	30	10^{-9}	
Recording studio	20		
Breathing	10		
	0†	10^{-12}	Inaudible

* The relationships illustrated in this table are necessarily approximate because the conditions of measurement are not defined. Typical levels should, however, be within about 10 dB of the stated values.

† O-dB sound pressure level (SPL) represents a reference sound pressure of 0.0002 μbar, or 0.00002 N/m^2.

The relationship between decibels and a selection of power and pressure ratios is given in Table 1.1.1. The footnote to the table describes a simple process for interpolating between these values, an exercise that helps to develop a feel for the meaning of the quantities.

The representation of the relative magnitudes of sound pressures and powers in decibels is important, but there is no indication of the absolute magnitude of either quantity being compared. This limitation is easily overcome by the use of a universally accepted reference level with which others are compared. For convenience the standard reference level is close to the smallest sound that is audible to a person with normal hearing. This defines a scale of *sound pressure level* (SPL), in which 0 dB represents a sound level close to the hearing-threshold level for middle and high frequencies (the most sensitive range). The SPL of a sound therefore describes, in decibels, the relationship between the level of that sounds and the reference level. Table 1.1.2 gives examples of SPLs of some common sounds with the corresponding intensities and an indication of listener reactions. From this table it is clear that the musically useful range of SPLs extend from the level of background noises in quiet surroundings to levels at which listeners begin to experience auditory discomfort and nonauditory sensations of feeling or pain in the ears themselves.

While some sound sources, such as chain saws and power mowers, produce a relatively constant sound output, others, like a 75-piece orchestra, are variable. The sound from such an orchestra might have a *peak factor* of 20 to 30 dB; the momentary, or peak, levels can be this amount higher than the long-term average SPL indicated [4].

The sound power produced by sources gives another perspective on the quantities being described. In spite of some impressively large sounds, a full symphony orchestra produces only about 1 acoustic watt when working through a typical musical passage. On crescendos with percussion, though, the levels can be of the order of 100 W. A bass drum alone can produce about 25 W of acoustic power of peaks. All these levels are dependent on the instruments and how they are played. Maximum sound output from cymbals might be 10 W; from a trombone, 6 W; and from a piano, 0.4 W [5]. By comparison, average speech generates about 25 µW, and a present-day jet liner at takeoff between 50 and 100 kW. Small gasoline engines produce from 0.001 to 1.0 acoustic watt, and electric home appliances less than 0.01 W [6].

1.1.4 References

1. Beranek, Leo L: *Acoustics,* McGraw-Hill, New York, N.Y., 1954.
2. Wong, G. S. K.: "Speed of Sound in Standard Air," *J. Acoust. Soc. Am*., vol. 79, pp. 1359–1366, 1986.
3. Pickett, J. M.: *The Sounds of Speech Communications*, University Park Press, Baltimore, MD, 1980.
4. Ehara, Shiro: "Instantaneous Pressure Distributions of Orchestra Sounds," *J. Acoust. Soc. Japan*, vol. 22, pp. 276–289, 1966.
5. Stephens, R. W. B., and A. E. Bate: *Acoustics and Vibrational Physics,* 2nd ed., E. Arnold (ed.), London, 1966.
6. Shaw, E. A. G.: "Noise Pollution—What Can be Done?" *Phys. Today*, vol. 28, no. 1, pp. 46–58, 1975.

Chapter

1.2

Sound Propagation

Floyd E. Toole

E. A. G. Shaw, G. A. Daigle, M. R. Stinson

1.2.1 Introduction

Sound propagating away from a source diminishes in strength at a rate determined by a variety of circumstances. It also encounters situations that can cause changes in amplitude and direction. Simple reflection is the most obvious process for directional change, but with sound there are also some less obvious mechanisms.

1.2.2 Inverse-Square and Other Laws

At increasing distances from a source of sound the level is expected to decrease. The rate at which it decreases is dictated by the directional properties of the source and the environment into which it radiates. In the case of a source of sound that is small compared with the wavelength of the sound being radiated, a condition that includes many common situations, the sound spreads outward as a sphere of ever-increasing radius. The sound energy from the source is distributed uniformly over the surface of the sphere, meaning that the intensity is the sound power output divided by the surface area at any radial distance from the source. Because the area of a sphere is $4\pi r^2$, the relationship between the sound intensities at two different distances is

$$\frac{I_1}{I_2} = \frac{r_2^2}{r_1^2} \tag{1.2.1}$$

where I_1 = intensity at radius r_1, I_2 = intensity at radius r_2, and

$$\text{Level difference} = 10 \log \frac{r_2^2}{r_1^2} = 20 \log \frac{r_2}{r_1} \text{ dB} \tag{1.2.2}$$

This translates into a change in sound level of 6 dB for each doubling or halving of distance, a convenient mnemonic.

In practice, however, this relationship must be used with caution because of the constraints of real environments. For example, over long distances outdoors the absorption of sound by the ground and the air can modify the predictions of simple theory [1]. Indoors, reflected sounds can sustain sound levels to greater distances than predicted, although the estimate is correct over moderate distances for the *direct sound* (the part of the sound that travels directly from source to receiver without reflection). Large sound sources present special problems because the sound waves need a certain distance to form into an orderly wave-front combining the inputs from various parts of the source. In this case, measurements in what is called the *near field* may not be representative of the integrated output from the source, and extrapolations to greater distances will contain errors. In fact, the *far field* of a source is sometimes defined as being distances at which the inverse-square law holds true. In general, the far field is where the distance from the source is at least 2 to 3 times the distance between the most widely separated parts of the sound source that are radiating energy at the same frequency.

If the sound source is not small compared with the wavelength of the radiated sound, the sound will not expand outward with a spherical wavefront and the rate at which the sound level reduces with distance will not obey the inverse-square law. For example, a sound source in the form of a line, such as a long column of loudspeakers or a long line of traffic on a highway, generates sound waves that expand outward with a cylindrical wavefront. In the idealized case, such sounds attenuate at the rate of 3 dB for each doubling of distance.

1.2.3 Sound Reflection and Absorption

A sound source suspended in midair radiates into a *free field* because there is no impediment to the progress of the sound waves as they radiate in any direction. The closest indoor equivalent of this is an *anechoic room*, in which all the room boundaries are acoustically treated to be highly absorbing, thus preventing sounds from being reflected back into the room. It is common to speak of such situations as sound propagation in *full space*, or 4π *steradians* (sr; the units by which solid angles are measured).

In normal environments sound waves run into obstacles, such as walls, and the direction of their propagation is changed. Figure 1.2.1 shows the *reflection* of sound from various surfaces. In this diagram the pressure crests of the sound waves are represented by the curved lines, spaced one wavelength apart. The radial lines show the direction of sound propagation and are known as *sound rays*. For reflecting surfaces that are large compared with the sound wavelength, the normal *law of reflection* applies: the angle that the incident sound ray makes with the reflecting surface equals the angle made by the reflected sound ray.

This law also holds if the reflecting surface has irregularities that are small compared with the wavelength, as shown in Figure 1.2.1*c*, where it is seen that the irregularities have negligible effect. If, however, the surface features have dimensions similar to the wavelength of the incident sound, the reflections are *scattered* in all directions. At wavelengths that are small compared with

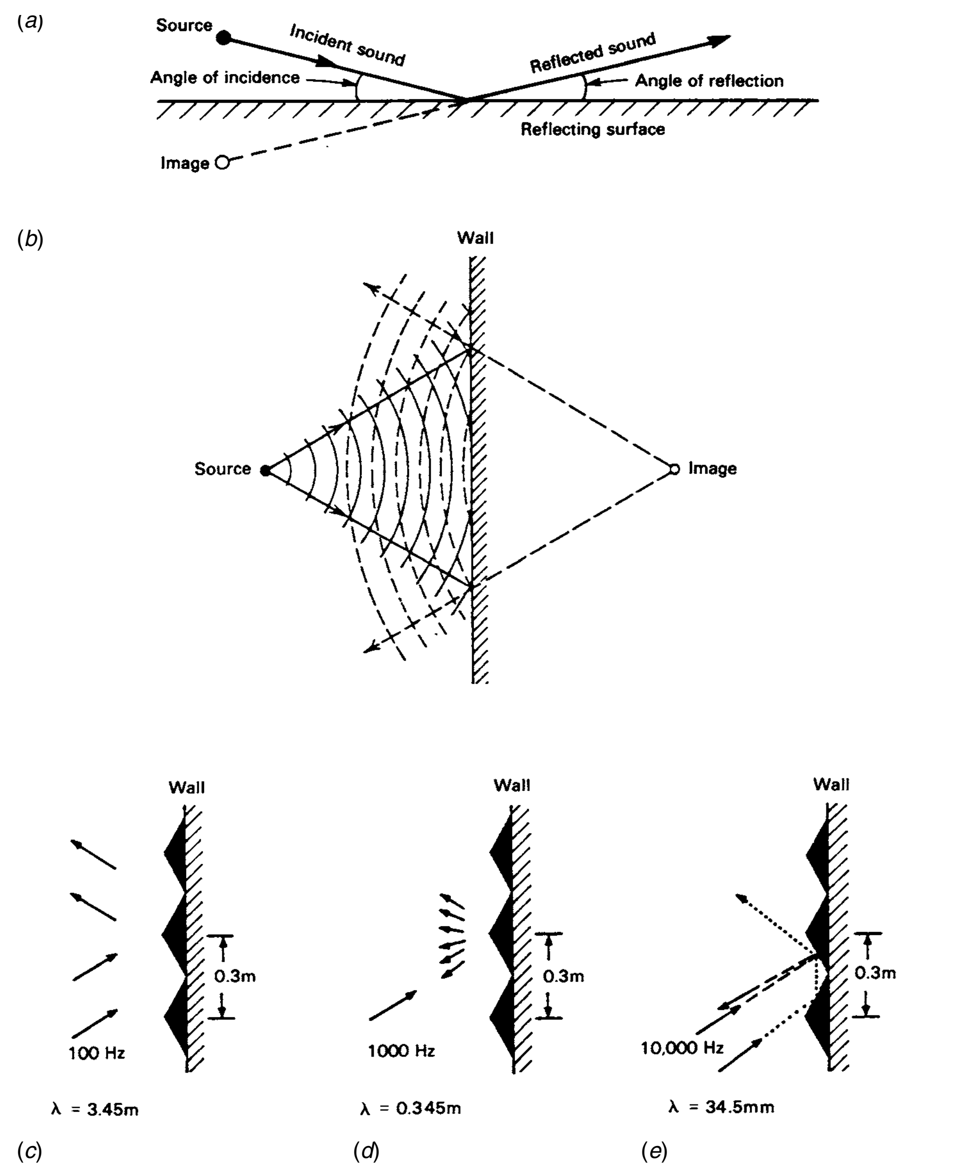


Figure 1.2.1 (*a*) The relationship between the incident sound, the reflected sound, and a flat reflecting surface, illustrating the law of reflection. (*b*) A more elaborate version of (*a*), showing the progression of wavefronts (the curved lines) in addition to the sound rays (arrowed lines). (*c*) The reflection of sound having a frequency of 100 Hz (wavelength 3.45 m) from a surface with irregularities that are small compared with the wavelength. (*d*) When the wavelength of the sound is similar to the dimensions of the irregularities, the sound is scattered in all directions. (*e*) When the wavelength of the sound is small compared with the dimensions of the irregularities, the law of reflection applies to the detailed interactions with the surface features.

the dimensions of the surface irregularities, the sound is also sent off in many directions but, in this case, as determined by the rule of reflections applied to the geometry of the irregularities themselves.

If there is perfect reflection of the sound, the reflected sound can be visualized as having originated at an image of the real source located behind the reflector and emitting the same sound power. In practice, however, some of the incident sound energy is *absorbed* by the reflecting surface; this fraction is called the *sound absorption coefficient* of the surface material. A coefficient of 0.0 indicates a perfect reflector, and a coefficient of 1.0 a perfect absorber; intermediate values indicate the portion of the incident sound energy that is dissipated in the surface and is not reflected. In general, the sound absorption coefficient for a material is dependent on the frequency and the angle of incidence of the sound. For simplicity, published values are normally given for octave bands of frequencies and for random angles of incidence.

1.2.3a Interference: The Sum of Multiple Sound Sources

The principle of *superposition* states that multiple sound waves (or electrical signals) appearing at the same point will add linearly. Consider two sound waves of identical frequency and amplitude arriving at a point in space from different directions. If the waveforms are exactly in step with each other, i.e., there is no phase difference, they will add perfectly and the result will be an identical waveform with double the amplitude of each incoming sound (6-dB-higher SPL). Such *in-phase* signals produce *constructive interference*. If the waveforms are shifted by one-half wavelength (180° phase difference) with respect to each other, they are *out of phase*; if the pressure fluctuations are precisely equal and opposite, *destructive interference* occurs, and perfect cancellation results.

In practice, interference occurs routinely as a consequence of direct and reflected sounds adding at a microphone or a listener's ear. The amplitude of the reflected sound is reduced because of energy lost to absorption at the reflecting surface and because of inverse-square-law reduction related to the additional distance traveled. This means that constructive interference yields sound levels that are increased by less than 6 dB and that destructive interference results in imperfect cancellations that leave a residual sound level. Whether the interference is constructive or destructive depends on the relationship between the extra distance traveled by the reflection and the wavelength of the sound.

Figure 1.2.2 shows the direct and reflected sound paths for an omnidirectional source and receivers interacting with a reflecting plane. Note that there is an acoustically mirrored source, just as there would be a visually mirrored one if the plane were optically reflecting. If the distance traveled by the direct sound and that traveled by the reflected sound are different by an amount that is small and is also small compared with a wavelength of the sound under consideration (receiver R_1), the interference at the receiver will be constructive. If the plane is perfectly reflecting, the sound at the receiver will be the sum of two essentially identical sounds and the SPL will be about 6 dB higher than the direct sound alone. Constructive interference will also occur when the difference between the distances is an even multiple of half wavelengths. Destructive interference will occur for odd multiples of half wavelengths.

As the path length difference increases, or if there is absorption at the reflective surface, the difference in the sound levels of the direct and reflected sounds increases. For receivers R_2 and R_3 in Figure 1.2.2, the situation will differ from that just described only in that, because of the

(a)

(b)

(c)

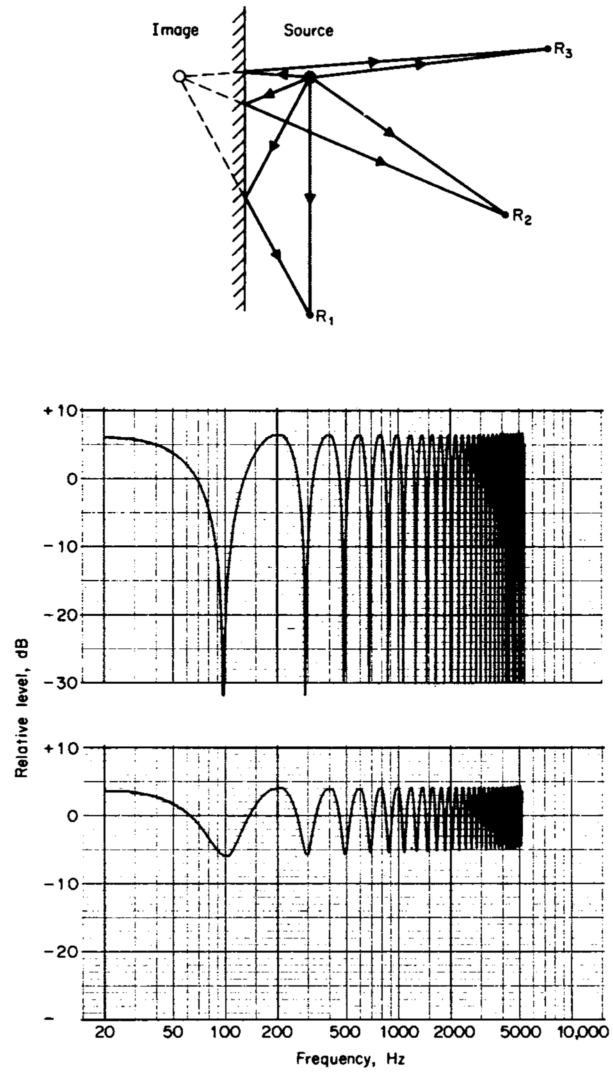


Figure 1.2.2 (*a*) Differing direct and reflected path lengths as a function of receiver location. (*b*) The interference pattern resulting when two sounds, each at the same sound level (0 dB) are summed with a time delay of just over 5 ms (a path length difference of approximately 1.7 m). (*c*) The reflection signal has been attenuated by 6 dB (it is now at a relative level of –6 dB, while the direct sounds remains at 0 dB); the maximum sound level is reduced, and perfect nulls are no longer possible. The familiar comb-filtering pattern remains.

additional attenuation of the reflected signal, the constructive peaks will be significantly less than 6 dB and the destructive dips will be less than perfect cancellations.

For a fixed geometrical arrangement of source, reflector, and receiver, at sufficiently low frequencies the direct and reflected sounds add. As the wavelength is reduced (frequency rising), the sound level at the receiver will decline from the maximum level in the approach to the first destructive interference at $\lambda/2 = r_2 - r_1$, where the level drops to a null. Continuing upward in frequency, the sound level at the receiver rises to the original level when $\lambda = r_2 - r_1$, falls to another null at $3\lambda/2 = r_2 - r_1$, rises again at $2\lambda = r_2 - r_1$, and so on, alternating between maxima and minima at regular intervals in the frequency domain. The plot of the frequency response of such a transmission path is called an *interference pattern*. It has the visual appearance of a comb, and the phenomenon has also come to be called *comb filtering* (see Figure 1.2.2*b*).

Taking a more general view and considering the effects averaged over a range of frequencies, it is possible to generalize as follows for the influence of a single reflecting surface on the sound level due to the direct sound alone [2].

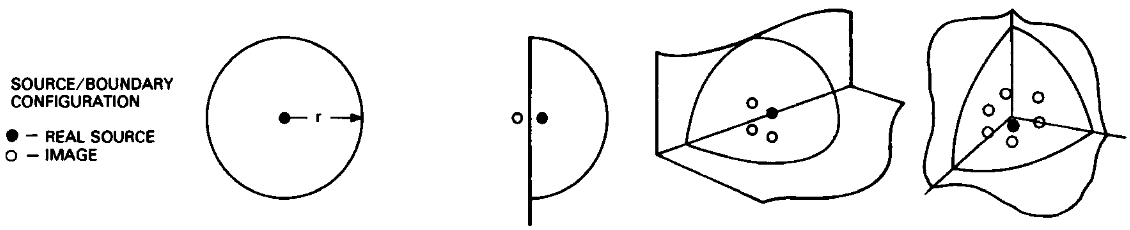
- When $r_2 - r_1$ is much less than a wavelength, the sound level at the receiver will be elevated by 6 dB or less, depending on the surface absorption and distances involved.
- When $r_2 - r_1$ is approximately equal to a wavelength, the sound level at the receiver will be elevated between 3 and 6 dB, depending on the specific circumstances.
- When $r_2 - r_1$ is much greater than a wavelength, the sound level at the receiver will be elevated by between 0 and 3 dB, depending on the surface absorption and distances involved.

A special case occurs when the sound source, such as a loudspeaker, is mounted in the reflecting plane itself. There is no path length difference, and the source radiates into a hemisphere of free space, more commonly called a *half space,* or 2π sr. The sound level at the receiver is then elevated by 6 dB at frequencies where the sound source is truly omnidirectional, which—in practice—is only at low frequencies.

Other reflecting surfaces contribute additively to the elevation of the sound level at the receiver in amounts that can be arrived at by independent analysis of each. Consider the situation in which a simple point monopole (omnidirectional) source of sound is progressively constrained by reflecting planes intersecting at right angles. In practice this could be the boundaries of a room that are immediately adjacent to a loudspeaker which, at very low frequencies, is effectively an omnidirectional source of sound. Figure 1.2.3 summarizes the relationships between four common circumstances, where the sound output from the source radiates into solid angles that reduce in stages by a factor of 2. These correspond to a loudspeaker radiating into free space (4π sr), placed against a large reflecting surface (2π sr), placed at the intersection of two reflecting surfaces (π sr), and placed at the intersection of three reflecting surfaces ($\pi/2$ sr). In all cases the dimensions of the source and its distance from any of the reflecting surfaces are assumed to be a small fraction of a wavelength. The source is also assumed to produce a constant volume velocity of sound output; i.e., the volumetric rate of air movement is constant throughout.

By using the principles outlined here and combining the outputs from the appropriate number of image sources that are acoustically mirrored in the reflective surfaces, it is found that the sound pressure at a given radius increases in inverse proportion to the reduction in solid angle; sound pressure increases by a factor of 2, or 6 dB, for each halving of the solid angle.

The corresponding sound intensity (the sound power passing through a unit surface area of a sphere of the given radius) is proportional to pressure squared. Sound intensity therefore



SOLID ANGLE SEEN BY SOURCE (steradians)	4π	2π	π	$\pi/2$
SOUND PRESSURE AT RADIUS r	p_r (0dB)	$2\ p_r$ (+ 6dB)	$4\ p_r$ (+ 12dB)	$8\ p_r$ (+ 18 dB)
SOUND INTENSITY AT RADIUS r	I_r (0dB)	$4I_r$ (+ 6dB)	$16I_r$ (+ 12dB)	$64I_r$ (+ 18dB)
SURFACE AREA AT RADIUS r	A_r	$\frac{A_r}{2}$	$\frac{A_r}{4}$	$\frac{A_r}{8}$
SOUND POWER RADIATED INTO THE SOLID ANGLE P = IA	$= I_r A_r$	$= 4\ I_r \frac{A_r}{2}$	$= 16\ I_r \frac{A_r}{4}$	$= 64\ I_r \frac{A_r}{8}$
	= Pr (0dB)	= 2 Pr (+ 3dB)	= 4 Pr (+ 6dB)	= 8 Pr (+ 9dB)

Figure 1.2.3 Behavior of a point monopole sound source in full space (4π) and in close proximity to reflecting surfaces that constrain the sound radiation to progressively smaller solid angeles. (*After* [3].)

increases by a factor of 4 for each halving of the solid angle. This also is 6 dB for each reduction in angle because the quantity is power rather than pressure.

Finally, multiplying the sound intensity by the surface area at the given radius yields the total sound power radiated into the solid angle. Because the surface area at each transition is reduced by a factor of 2, the total sound power radiated into the solid angle increases by a factor of 2, or 3 dB, for each halving of the solid angle.

By applying the reverse logic, reducing the solid angle by half increases the rate of energy flow into the solid angle by a factor of 2. At a given radius, this energy flows through half of the surface area that it previously did, so that the sound intensity is increased by a factor of 4; i.e., pressure squared is increased by a factor of 4. This means that sound pressure at that same radius is increased by a factor of 2.

The simplicity of this argument applies when the surfaces shown in Figure 1.2.3 are the only ones present; this can only happen outdoors. In rooms there are the other boundaries to consider, and the predictions discussed here will be modified by the reflections, absorption, and standing-wave patterns therein.

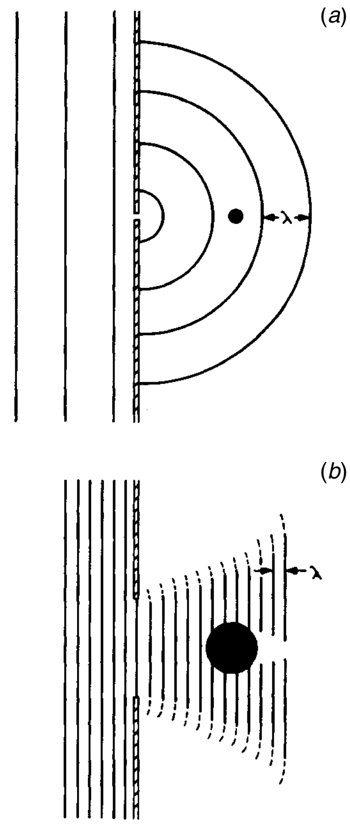


Figure 1.2.4 Stylized illustration of the diffraction of sound waves passing through openings and around obstacles. (*a*) The case where the wavelength is large compared with the size of the opening and the obstacle. (*b*) The case where the wavelength is small compared with the size of the opening and the obstacle.

1.2.3b Diffraction

The leakage of sound energy around the edges of an opening or around the corners of an obstacle results in a bending of the sound rays and a distortion of the wave-front. The effect is called *diffraction*. Because of diffraction it is possible to hear sounds around corners and behind walls—anywhere there might have been an "acoustical shadow." In fact, acoustical shadows exist, but to an extent that is dependent on the relationship between the wavelength and the dimensions of the objects in the path of the sound waves.

When the openings or obstructions are small compared with the wavelength of the sound, the waves tend to spread in all directions and the shadowing effect is small. At higher frequencies, when the openings or obstructions are large compared with the wavelengths, the sound waves tend to continue in their original direction of travel and there is significant shadowing. Figure 1.2.4 illustrates the effect.

The principle is maintained if the openings are considered to be the diaphragms of loudspeakers. If one wishes to maintain wide dispersion at all frequencies, the radiating areas of the driver units must progressively reduce at higher frequencies. Conversely, large radiating areas can be used to restrict the dispersion, though the dimensions required may become impractically large at

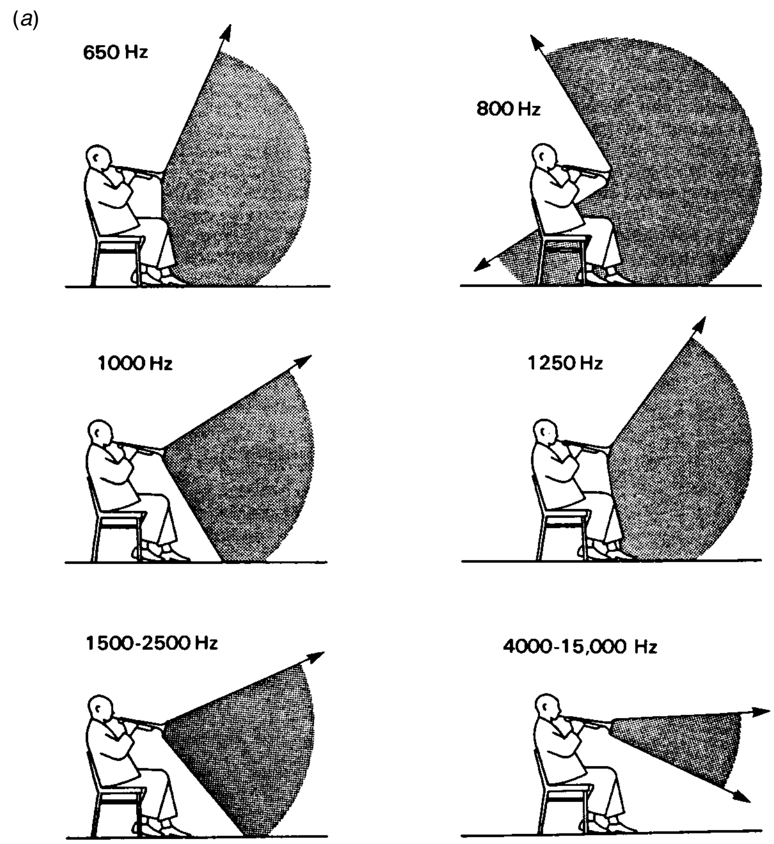


Figure 1.2.5 A simplified display of the main sound radiation directions at selected frequencies for: (*a*) a trumpet, (*b*, next page) a cello. (*From* [4]. *Used with permission.*)

low frequencies. As a consequence, most loudspeakers are approximately omnidirectional at low frequencies.

Sounds radiated by musical instruments obey the same laws. Low-frequency sounds from most instruments and the human voice radiate in all directions. Higher-frequency components can exhibit quite strong directional biases that are dependent on the size and orientation of the major sound-radiating elements. Figure 1.2.5*a* shows the frequency-dependent directivities of a trumpet, a relatively simple source. Compare this with the complexity of the directional characteristics of a cello (Figure 1.2.5*b*). It is clear that no single direction is representative of the total sound output from complex sound sources—a particular difficulty when it comes to choosing microphone locations for sound recordings. Listeners at a live performance hear a combination of all the directional components as spatially integrated by the stage enclosure and the hall itself.

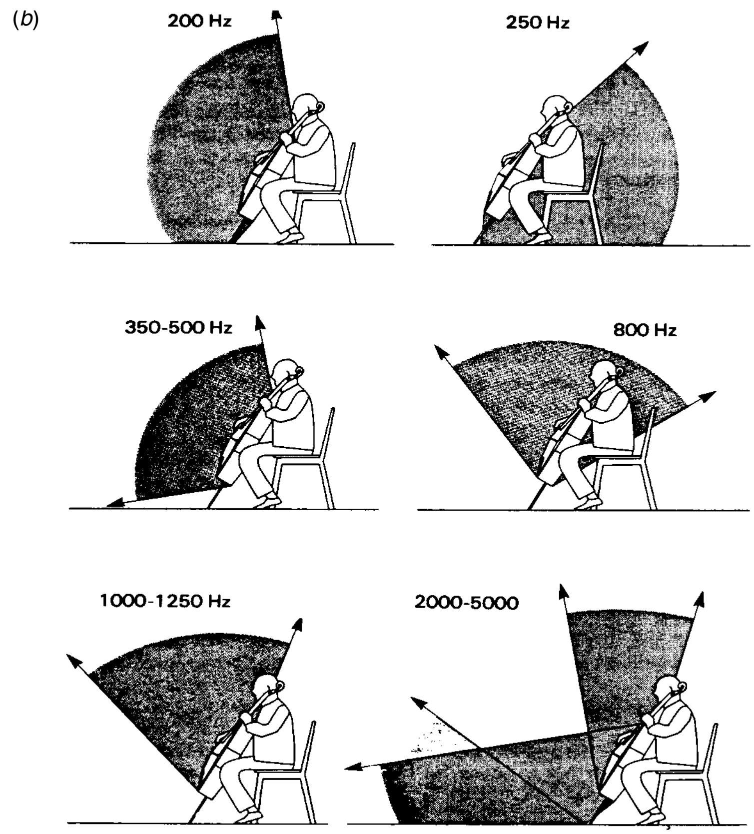


Figure 1.2.5*b*

1.2.3c Refraction

Sound travels faster in warm air than in cold and faster downwind than upwind. These factors can cause sound rays to be bent, or *refracted*, when propagating over long distances in vertical gradients of wind or temperature. Figure 1.2.6 shows the downward refraction of sound when the propagation is downwind or in a *temperature inversion*, as occurs at night when the temperature near the ground is cooler than the air higher up. Upward refraction occurs when the propagation is upwind or in a *temperature lapse*, a typical daytime condition when the air temperature falls with increasing altitude. Thus, the ability to hear sounds over long distances is a function of local climatic conditions; the success of outdoor sound events can be significantly affected by the time of day and the direction of prevailing winds.

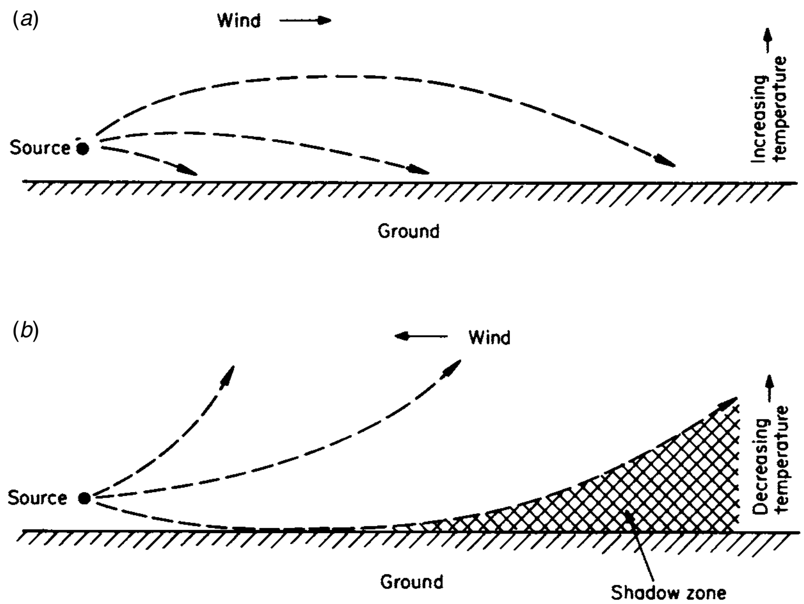


Figure 1.2.6 The refraction of sound by wind and by temperature gradients: (*a*) downwind or in a temperature inversion, (*b*) upwind or in a temperature lapse. (*From* [1]. *Used with permission.*)

1.2.4 References

1. Piercy, J. E., and T. F. W. Embleton: "Sound Propagation in the Open Air," in *Handbook* of *Noise Control*, 2d ed., C. M. Harris (ed.), McGraw-Hill, New York, N.Y., 1979.
2. Waterhouse, R. V., and C. M. Harris: "Sound in Enclosed Spaces," in *Handbook of Noise Control*, 2d ed., C. M. Harris (ed.), McGraw-Hill, New York, N.Y., 1979.
3. Olson, Harry F.: *Acoustical Engineering*, Van Nostrand, New York, N.Y., 1957.
4. Meyer, J.: *Acoustics and the Performance of Music*, Verlag das Musikinstrument, Frankfurt am Main, 1987.

Chapter

1.3

Resonance

Floyd E. Toole

E. A. G. Shaw, G. A. Daigle, M. R. Stinson

1.3.1 Introduction

A vibrating system of any kind that is driven by and is completely under the control of an external source of energy is in a state of *forced vibration*. The activity within such a system after the external force has been removed is known as *free vibration*. In this condition most systems exhibit a tendency to move at a natural or *resonant* frequency, declining with time at a rate determined by the amount of energy dissipation, or *damping*, in the system. The resonances in some musical instruments have little damping, as the devices are intended to resonate and produce sound at specific frequencies in response to inputs, such as impacts or turbulent airflow, that do not have any specific frequency characteristics. Most instruments provide the musician with some control over the damping so that the duration of the notes can be varied.

1.3.2 Fundamental Properties

If the frequency of the driving force is matched to the natural frequency of the resonant system, the magnitude of the vibration and the efficiency of the energy transfer are maximized.

These and other points are illustrated in Figure 1.3.1, which shows three versions of a resonant system having different amounts of damping. The term commonly used to describe this characteristic of resonant systems is the *quality factor*, *Q*, a measure of the lightness of damping in a system. The system in Figure 1.3.1*a* has a *Q* of 1; it is well damped. The system in Figure 1.3.1*b* is less welt damped and has a *Q* of 10, while that in Figure 1.3.1*c* has little damping and is described as having a *Q* of 50. As a practical example, the resonance of a loudspeaker in an enclosure would typically have a *Q* of 1 or less. Panel resonances in enclosures might have *Q*s in the region of 10 or so. Resonances with a *Q* of 50 or more would be rare in sound reproducers but common in musical instruments.

On the left in Figure 1.3.1 can be seen the behavior of these systems when they are forced into oscillation by a pure tone tuned to the resonance frequency of the systems, 1000 Hz. When the

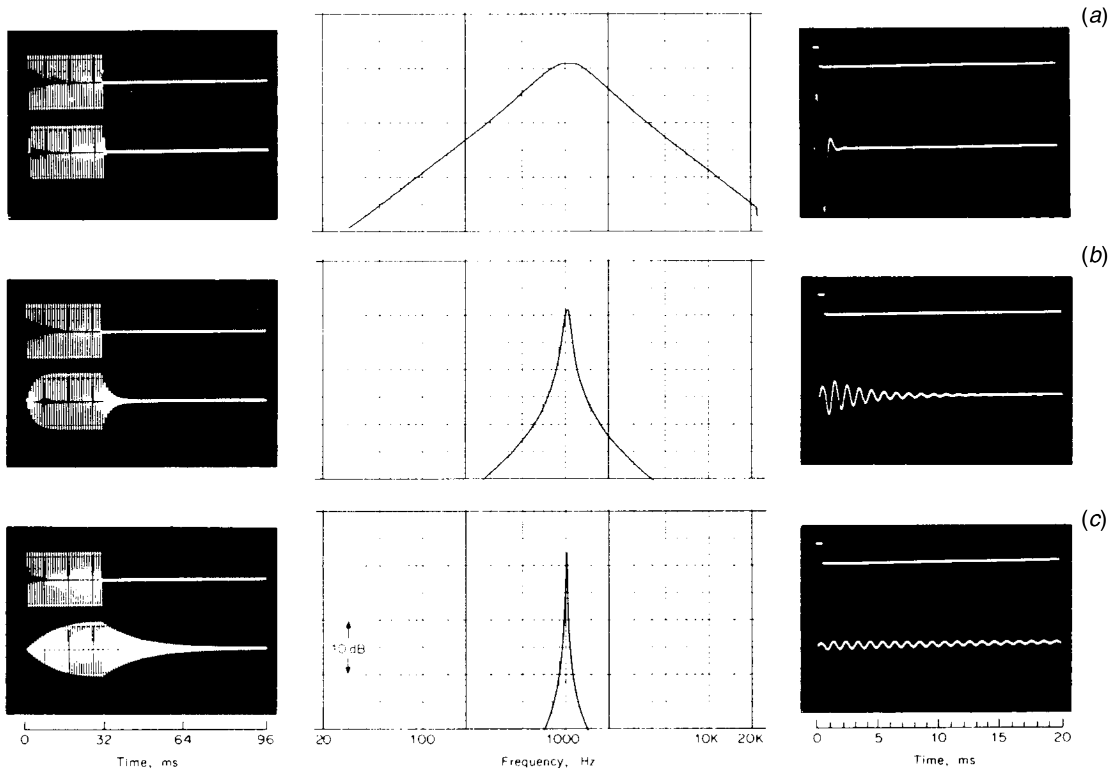


Figure 1.3.1 The frequency responses of three resonant systems and their behavior in conditions of forced and free vibration. The system show in (*a*) has the least damping ($Q = 1$), system (*b*) has moderate damping ($Q = 10$), and the system shown in (*c*) has the least damping ($Q = 50$).

tone is turned on and off, the systems respond with a speed that is in inverse proportion to the Q. The low-Q resonance responds quickly to the onset of the tone and terminates its activity with equal brevity. The medium-Q system responds at a more leisurely rate and lets the activity decay at a similar rate after the cessation of the driving signal. The high-Q system is slow to respond to the driving signal and sustains the activity for some time after the interval of forced oscillation.

In the preceding example the forcing signal was optimized in frequency, in that it matched the resonance frequency of the system, and it was sustained long enough for the system to reach its level of maximum response. On the right of Figure 1.3.1 are shown the responses of these systems to an impulse signal brief in the time domain but having energy over a wide range of frequencies including that of the resonant system. In Figure 1.3.1*a* the low-Q system is shown responding energetically to this signal but demonstrating little sustained activity. In Figure 1.3.1*b* and 1.3.1*c* the higher-Q systems respond with progressively reduced amplitude but with progressively sustained ringing alter the pulse has ended. Note that the ringing is recognizably at the resonance frequency, 1 cycle/ms.

In the center of Figure 1.3.1 are shown the amplitude-frequency responses or, more commonly, the *frequency responses* of the systems. These curves show the amplitude of system response when the frequency of a constant driving signal is varied from well below the resonance frequency to well above it. The low-Q system Figure 1.3.1*a* is seen to respond to signals over a wide frequency range, but the higher-Q systems become progressively more frequency-selective.

In this illustration, the maximum amplitudes of the system responses at resonance were adjusted to be approximately equal. Such is often the case in electronic resonators used in filters, frequency equalizers, synthesizers, and similar devices. In simple resonant systems in which everything else is held equal and only the damping is varied, the maximum amplitude response would be highest in the system with the least dissipation: the high-Q system, Figure 1.3.1*c.* Adding damping to the system would reduce the maximum amplitude, so that the system with the lowest Q, having the highest damping or losses, would respond to the widest range of frequencies, but with reduced amplitude [1].

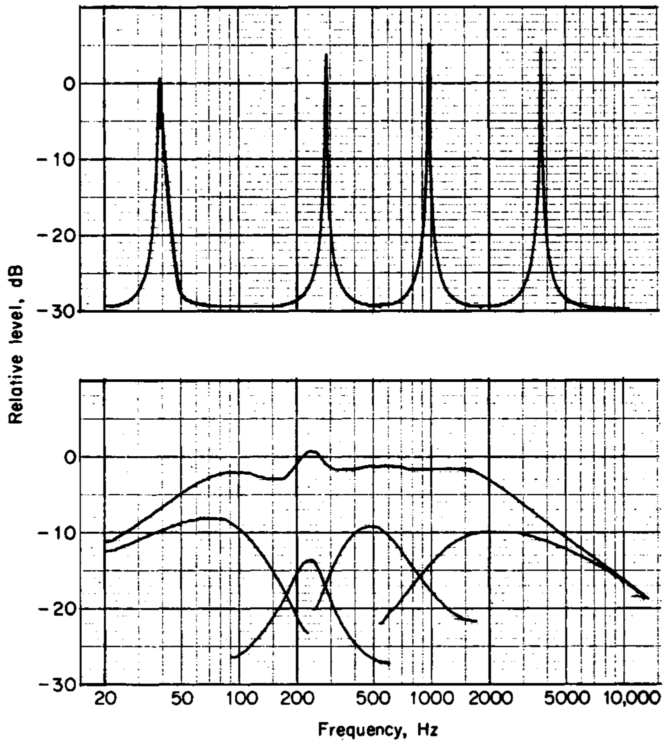
Figure 1.3.2 shows the frequency responses of two systems with multiple resonances. In 1.3.2*a* the resonances are such that they respond independently to driving forces at single frequencies. In 1.3.2*b* an input at any single frequency would cause some activity in all the resonators but at different amplitudes in each one. The series of high-Q resonators in Figure 1.3.2*a* is characteristic of musical instruments, where the purpose is the efficient production of sound at highly specific frequencies. The overlapping set of low-Q resonators in Figure 1.3.2*b* are the filters of a parametric equalizer in which the frequency, Q, and amplitude of the filters are individually adjustable to provide a variable overall frequency response for a sound-recording or sound-reproducing system.

A special case of Figure 1.3.2*b* would be a multiway loudspeaker system intended for the reproduction of sounds of all kinds. In this case, the selection of loudspeaker units and their associated filters (crossovers) would be such that, in combination, they resulted in an overall amplitude response that is flat (the same at all frequencies) over the required frequency range. Such a system would be capable of accurately recreating any signal spectrum. For the loudspeaker or any system of multiple filters or resonant elements to accurately pass or reproduce a complex waveform, there must be no phase shift at the important frequencies. In technical terms this would be assessed by the *phase-frequency response*, or *phase response,* of the system showing the amount of phase shift at frequencies within the system bandwidth.

Resonant systems can take any of several forms of electrical, mechanical, or acoustical elements or combinations thereof. In electronics, resonators are the basis for frequency-selective or tuned circuits of all kinds, from radios to equalizers and music synthesizers. Mechanical resonances are the essential pitch determinants of tuning forks, bells, xylophones, and glockenspiels. Acoustical resonances are the essential tuning devices of organs and other wind instruments. Stringed instruments involve combinations of mechanical and acoustical resonances in the generation and processing of their sounds, as do reed instruments and the human voice.

The voice is a good example of a complex resonant system. The sound originates as a train of pulses emanating from the voice box. This excites a set of resonances in the vocal tract so that the sound output from the mouth is emphasized at certain frequencies. In spectral terms, the envelope of the line spectrum is modified by the frequency response of the resonators in the vocal tract. These resonances are called *formants,* and their frequencies contribute to the individual character of voices. The relative amplitudes of the resonances are altered by changing the physical form of the vocal tract so as to create different vowel sounds, as illustrated in Figure 1.3.3 [2–4].

(*a*)



(*b*)

Figure 1.3.2 Two systems with multiple resonances: (*a*) well-separated high-*Q* resonances that can respond nearly independently of each other, as in the notes of a musical instrument; (*b*) the four filters of a parametric equalizer designed to produce overlapping low-*Q* resonance curves (bottom traces) which are combined to produce a total response (top trace) that may bear little resemblance to the individual contributions.

1.3.2a Resonance in Pipes

When the diameter of a pipe is small compared with the wavelength, sound will travel as plane waves perpendicular to the length of the pipe. At a closed end the sound is reflected back down the pipe in the reverse direction. At an open end, some of the sound radiates outward and the remainder is reflected backward, but with a pressure reversal (180° phase shift). The pressure distribution along the pipe is therefore the sum of several sound waves traveling backward and forward. At most frequencies the summation of these several waves results in varying degrees of destructive interference, but at some specific frequencies the interference is only constructive and a pattern stabilizes in the form of *standing waves*. At these frequencies, the wavelengths of the sounds are such that specific end conditions of the tube are simultaneously met by the waves traveling in both directions, the sounds reinforce each other, and a resonant condition exists.

Figures 1.3.4 and 1.3.5 show the first three *resonant modes* for pipes open at both ends and for those with one end closed. The open ends prevent the pressures from building up, but the par-

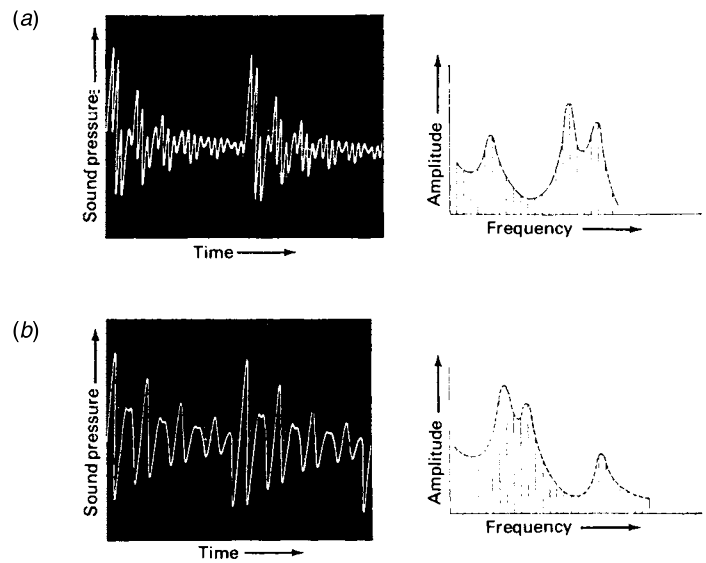


Figure 1.3.3 The waveforms and corresponding amplitude-frequency spectra of the vowel sounds "uh" (*a*) and "ah" (*b*). (*From* [3]. *Used with permission.*)

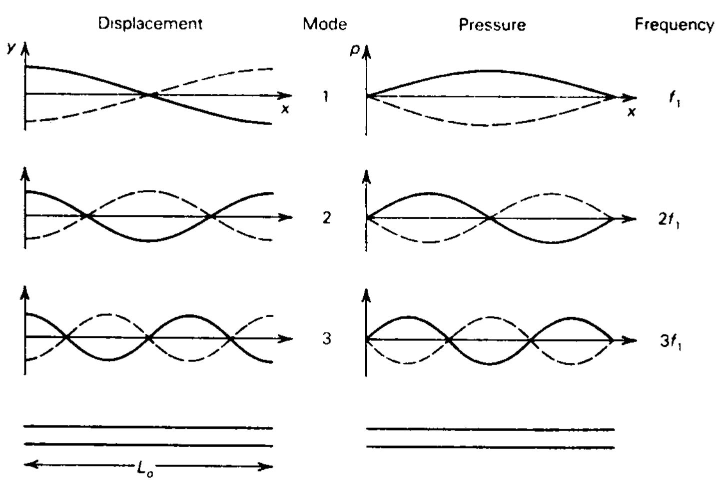


Figure 1.3.4 The first three resonant modes of air in a tube open at both ends. On the left are the patterns of particle displacement along the tube, showing the antinodes at the ends of the tube. At the right are the corresponding patterns of pressure, with the required nodes at the ends. The fundamental frequency is $c/2L_o$. (*From* [7]. *Used with permission.*)

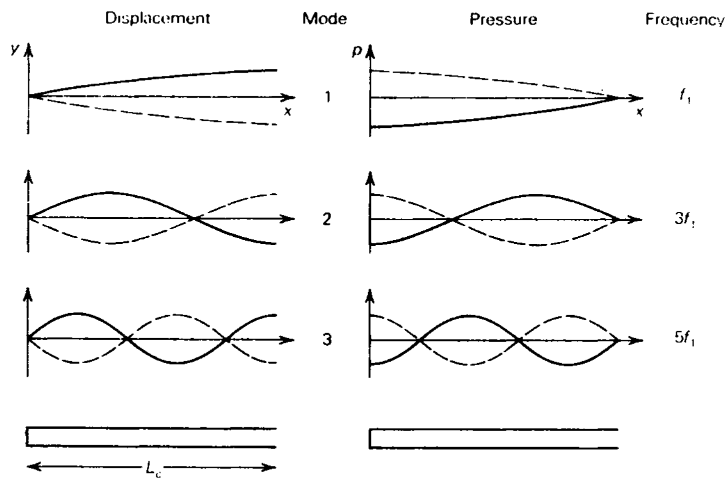


Figure 1.3.5 The first three resonant modes of air in a tube closed at one end. On the left are the patterns of particle displacement along the tube, and on the right are the pressure distributions. The fundamental frequency is $c/4L_o$. (*From* [7]. *Used with permission.*)

ticle displacements are unimpeded; the end condition for resonance is therefore a displacement maximum (*antinode*) and a pressure minimum (*node*) in the standing-wave pattern. A closed end does the reverse, forcing displacements to zero but permitting pressure to build up; the end condition for resonance is therefore a displacement node and a pressure antinode.

For a pipe open at both ends, the fundamental frequency has a wavelength that is double the length of the pipe; conversely, the pipe is one-half wavelength long. The fundamental frequency is therefore $f = c/2L_o$, where L_o is the length of the pipe in meters and c is the speed of sound: 345 m/s. Other resonances occur at all harmonically related frequencies: $2f_1$, $3f_1$, and so on.

A pipe closed at one end is one-quarter wavelength long at the fundamental resonance frequency; thus $f = c/4L_c$. In this case, however, the other resonances occur at odd harmonics only: $3f_1$, $5f_1$, and so on. A very simplistic view of the vocal tract considers it as a pipe, closed at the vocal cords, open at the mouth, and 175 mm long [4]. This yields a fundamental frequency of about 500 Hz and harmonics at 1500, 2500, and 3500 Hz. These are close to the formant frequencies appearing as resonance humps in the spectra of Figure 1.3.3.

Organ pipes are of both forms, although the pipes open at both ends produce the musically richer sound. To save space, pipes closed at one end are sometimes used for the lowest notes; these need be only one-fourth wavelength long, but they produce only odd harmonics.

In practice this simple theory must be modified slightly to account for what is called the *end correction*. This can be interpreted as the distance beyond the open end of the pipe over which the plane waves traveling down the pipe make the transition to spherical wavefronts as they diverge after passing beyond the constraints of the pipe walls. The pipe behaves as if is longer than its physical length by an amount equal to 0.62 times its radius. If the pipe has a flange or opens onto a flat surface, the end correction is 0.82 times the radius.

1.3.2b Resonance in Rooms and Large Enclosures

Sounds propagating in rectangular rooms and large enclosures are subject to standing waves between the reflecting boundaries. In taking a one-dimensional view for illustration, sounds reflecting back and forth between two parallel surfaces form standing waves at frequencies satisfying the boundary conditions requiring pressure antinodes and particle displacement nodes at the reflecting surfaces. The fundamental resonance frequency is that at which the separation is one-half wavelength. Other resonances occur at harmonics of this frequency. This same phenomenon exists between all opposing pairs of parallel surfaces, establishing three sets of resonances, dependent on the length, width, and height, known as the *axial modes* of the enclosure. Other resonances are associated with sounds reflected from four surfaces and propagating in a plane parallel to the remaining two. For example, sound can be reflected from the four walls and travel parallel to the floor and ceiling. The three sets of these resonances are called *tangential modes*. Finally, there are resonances involving sounds reflected from all surfaces in the enclosure, called *oblique modes*. All these resonant modes, or *eigentones*, can be calculated from

$$f_n = \frac{c}{2}\sqrt{\left(\frac{n_x}{l_x}\right)^2 + \left(\frac{n_y}{l_y}\right)^2 + \left(\frac{n_z}{l_z}\right)^2} \tag{1.3.1}$$

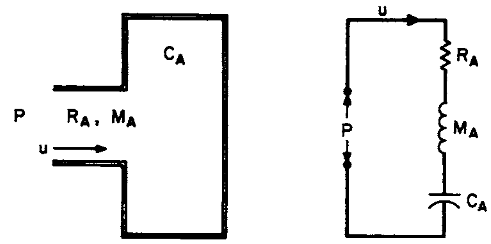
where:
f_n = frequency of the *n*th mode
n_x , n_y, n_z = integers with independently chosen values between 0 and ∞
l_x , l_y, l_z = dimensions of enclosure, m (ft)
c = speed of sound, m/s (ft/s)

It is customary to identify the individual modes by a combination of n_x , n_y, and n_z , as in (2, 0, 0), which identifies the mode as being the second-harmonic resonance along the *x* dimension of the enclosure. All axial modes are described by a single integer and two zeros. Tangential modes are identified by two integers and one zero, and oblique modes by three integers. The calculation of all modes for an enclosure would require the calculation of Equation (1.3.1) for all possible combinations of integers for n_x , n_y, and n_z .

The sound field inside an enclosure is therefore a complex combination of many modes, and after the sound input has been terminated, they can decay at quite different rates depending on the amount and distribution of acoustical absorption on the room boundaries. Because some energy is lost at every reflection, the modes that interact most frequently with the room boundaries will decay first. The oblique modes have the shortest average distance between reflections and are the first to decay, followed by the tangential modes and later by the axial modes. This means that the sound field in a room is very complex immediately following the cessation of sound production, and it rapidly deteriorates to a few energetic tangential and axial modes [5, 6].

The ratio of length to width to height of an enclosure determines the distribution of the resonant modes in the frequency domain. The dimensions themselves determine the frequencies of the modes. The efficiency with which the sound source and receiver couple to the various modes determines the relative influence of the modes in the transmission of sound from the source to the receiver. These factors are important in the design of enclosures for specific purposes. In a listening or control room, for example, the locations of the loudspeakers and listeners are largely determined by the geometrical requirements for good stereo listening and by restrictions

Figure 1.3.6 Physical representation of a Helmholtz resonator (left) and the corresponding symbolic representation as a series resonant acoustical circuit (right). Legend: P = sound pressure at mouth; μ = volume velocity at the port = particle velocity × port area; R_A = acoustical resistance; M_A = acoustical mass of the port; C_A = acoustical compliance of the volume.



imposed by the loudspeaker design. Accurate communication from the source to the receiver over a range of frequencies requires that the influential room modes be uniformly distributed in frequency. Clusters or gaps in the distribution of modes can cause sounds at some frequencies to be accentuated and others to be attenuated, altering the frequency response of the sound propagation path through the room. This causes the timbre of sounds propagated through the room to be changed.

Certain dimensional ratios have been promoted as having especially desirable mode distributions. Indeed, there are shapes like cubes and corridors that clearly present problems, but the selection of an ideal rectangular enclosure must accommodate the particular requirements of the application. Generalizations based on the simple application of Equation (1.3.1) assume that the boundaries of the enclosure are perfectly reflecting and flat, that all modes are equally energetic, and that the source and receiver are equally well coupled to them all. In practice it is highly improbable that these conditions will be met.

1.3.2c Resonance in Small Enclosures: Helmholtz Resonators

At frequencies where the wavelength is large compared with the interior dimensions of an enclosure, there is negligible wave motion because the sound pressure is nearly uniform throughout the volume. In these circumstances the *lumped-element* properties of the enclosed air dominate, and another form of resonance assumes control. Such *Helmholtz resonators* form an important class of acoustic resonators.

Figure 1.3.6 shows a simple cavity with a short ducted opening, like a bottle with a neck. Here the volume of air within the cavity acts as a spring for the mass of air in the neck, and the system behaves as the acoustical version of a mechanical spring-mass resonant system. It is also analogous to the electrical resonant circuit with elements as shown in the figure.

Acoustical compliance increases with the volume, meaning that the resonance frequency falls with increasing cavity volume. The acoustic mass (*inertance*) in the duct increases with the length of the duct and decreases with increasing duct area, leading to a resonance frequency that is proportional to the square root of the duct area and inversely proportional to the square root of the duct length.

Helmholtz resonators are the simplest form of resonating systems. They are found as the air resonance in the body of guitars, violins, and similar instruments, and they are the principal frequency-determining mechanism in whistles and ocarinas. They also describe the performance of loudspeaker-enclosure systems at low frequencies. The acoustical-mechanical-electrical analogs introduced here are the basis for the design of closed-box and reflex loudspeaker systems, resulting in closely predictable performance at low frequencies. At higher frequencies, standing waves form inside the box, and the tidy lumped-element concepts no longer apply.

1.3.2d Horns

If the open end of a tube has a diameter that is small compared with the wavelength of sound being propagated within it, most of the sound is reflected back into the tube, and if the wavelength is appropriate, standing waves result. At resonance, the acoustical activity is at its maximum, but the small tube opening is nevertheless a rather inefficient radiator of sound. If strong resonances are important and adequate input power is available, as in organ pipes, this is a desirable situation. Other devices, however, require the maintenance of strong standing waves, but with an improved radiation efficiency. With care this is achieved through the use of a flared development, or *horn*, at the end of the pipe. The shape and size of the horn determine, for every frequency, how much of the sound is reflected back into the tube and how much radiates outward.

The musical instruments of the brass family are all combinations of resonant pipes with a flaring bell at the output end. The shape of a trumpet bell, for example, is such that it has radiation efficiency that is low below about 1500 Hz and high above. By establishing strong resonances at the fundamental playing frequencies, the bell makes the instrument playable while imparting a bright sound character by efficiently radiating the higher harmonics of the basic pitch [7, 8].

On the other hand, a loudspeaker horn must have high radiation efficiency at all frequencies within its operating range; otherwise there will be resonances in a system that is intended to be free of such sources of tone color. The key to non-resonant behavior lies in the choice of flare shape and mouth size. The sound waves propagating outward must be allowed to expand at just the proper rate, maintaining close control over the directions of the particle velocities, so that the waves can emerge from the enlarged mouth with little energy reflected back to the loudspeaker. [5].

1.3.3 References

1. Main, Ian G.: *Vibrations and Waves in Physics,* Cambridge, London, 1978.
2. Pickett, J. M.: *The Sounds of Speech Communication,* University Park Press, Baltimore, MD, 1980.
3. Denes, Peter B., and E. N. Pinson: *The Speech Chain,* Bell Telephone Laboratories, Waverly, 1963.
4. Sundberg, Johan: “The Acoustics of the Singing Voice,” in *The Physics of Music,* introduction by C. M. Hutchins, Scientific American/Freeman, San Francisco, Calif., 1978.
5. Morse, Philip M.: *Vibrations and Sound,* 1964, reprinted by the Acoustical Society of America, New York, N.Y., 1976.
6. Mankovsky, V. S.: *Acoustics of Studios and Auditoria,* Focal Press, London, 1971.
7. Hall, Donald E.: *Musical Acoustics—An Introduction,* Wadsworth, Belmont, Calif, 1980.

8. Benade, A. H.: *Fundamentals of Musical Acoustics,* Oxford University Press, New York, N.Y., 1976.

Chapter

1.4

The Physical Nature of Hearing

Floyd E. Toole

E. A. G. Shaw, G. A. Daigle, M. R. Stinson

1.4.1 Introduction

The process of hearing begins with acoustical modifications to the sound waves as they interact with the head and the external ear, the visible portion of the system. These acoustical changes are followed by others in the ear canal and by a conversion of the sound pressure fluctuations into mechanical displacements by the eardrum. Transmitted through the mechanical coupling system of the middle ear to the inner ear, the displacement patterns are partially analyzed and then encoded in the form of neural signals. The signals from the two ears are cross-compared at several stages on the way to the auditory centers of the brain, where finally there is a transformation of the streams of data into perceptions of sound and acoustical space.

By these elaborate means we are able to render intelligible acoustical signals that, in technical terms, can be almost beyond description. In addition to the basic information, the hearing process keeps us constantly aware of spatial dimensions, where sounds are coming from, and the general size, shape, and decor of the space around us—a remarkable process indeed.

1.4.2 Anatomy of the Ear

Figure l.4.1*a* shows a cross section of the ear in a very simplified form in which the outer, middle, and inner ear are clearly identified. The head and the outer ear interact with the sound waves, providing acoustical amplification that is dependent on both direction and frequency, in much the same way as an antenna. At frequencies above about 2 kHz there are reflections and resonances in the complex folds of the *pinna* [1]. Consequently, sounds of some frequencies reach the *tympanic membrane* (eardrum) with greater amplitude than sounds of other frequencies. The amount of the sound pressure gain or loss depends on both the frequency and the angle of incidence of the incoming sound. Thus, the external ear is an important first step in the perceptual process, encoding sounds arriving from different directions with distinctive spectral characters. For example, the primary resonance of the external ear, at about 2.6 kHz, is most sensitive to

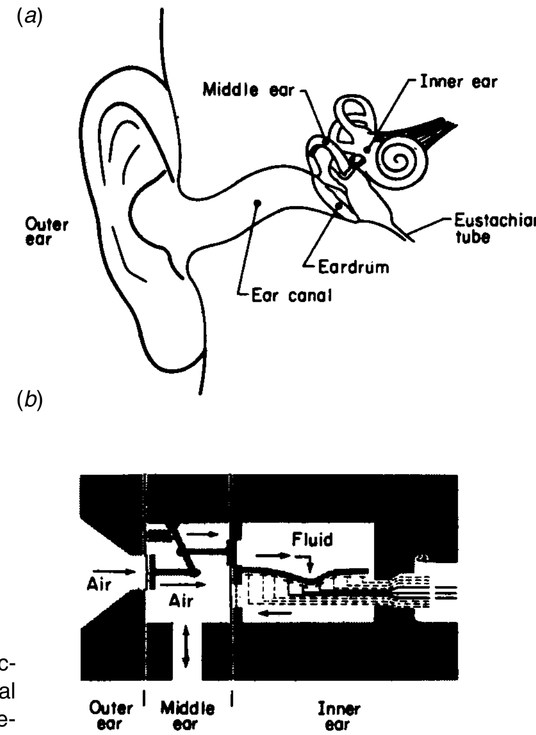


Figure 1.4.1 The human ear: (*a*) cross-sectional view showing the major anatomical elements, (*b*) a simplified functional representation.

sounds arriving from near 45° azimuth. This can be demonstrated by listening to a source of broadband sound while looking directly at it and then slowly rotating the head until one ear is pointing toward it. As the head is rotated through 45°, the sound should take on a "brighter" character as sounds in the upper midrange are accentuated. People with hearing problems use this feature of the ear to improve the intelligibility of speech when they unconsciously tilt the head, directing the ear toward the speaker. Continuing the rotation reveals a rapid dulling of the sound as the source moves behind the head. This is caused by acoustical shadowing due to diffraction by the pinna, a feature that helps to distinguish between front and back in sound localization.

At the eardrum the sound pressure fluctuations are transformed into movement that is coupled by means of the middle-ear bones (the *ossicular chain*) to the *oval window*, the input to the inner ear (*cochlea*). The middle ear increases the efficiency of sound energy transfer by providing a partial impedance match between sound in air, on the one hand, and wave motion in the liquid-filled inner ear, on the other. The inner ear performs the elaborate function of analyzing the sound into its constituent frequencies and converting the result into neural signals that pass up the auditory (eighth) nerve to the auditory cortex of the brain. From there sound is transformed into the many and varied perceptions that we take for granted. In the following discussions we shall be dealing with some of these functions in more detail.

1.4.3 Psychoacoustics and the Dimensions of Hearing

The physical dimensions of sound have parallels in the perceptual processes. The relationships are usually nonlinear, more complex than at first appearance, and somewhat variable among individuals as well as with time and experience. Nevertheless, they are the very essence of hearing.

The study of these relationships falls under the general umbrella of *psycho-acoustics*. A more specialized study, known as *psychophysics* or *psychometrics*, is concerned with quantification of the magnitudes of the sensation in relation to the magnitude of the corresponding physical stimulus.

1.4.3a Loudness

Loudness is the term used to describe the magnitude of an auditory sensation. It is primarily dependent upon the physical magnitude (sound pressure) of the sound producing the sensation, but many other factors are influential.

Sounds come in an infinite variety of frequencies, timbres, intensities, temporal patterns, and durations; each of these, as well as the characteristics of the individual listener and the context within which the sound is heard, has an influence on loudness. Consequently, it is impossible for a single graph or equation to accurately express the relationship between the physical quality and quantity of sound and the subjective impression of loudness. Our present knowledge of the phenomenon is incomplete, but there are some important experimentally determined relationships between loudness and certain measurable quantities of sound. Although it is common to present and discuss these relationships as matters of fact, it must always be remembered that they have been arrived at through the process of averaging the results of many experiments with many listeners. These are not precise engineering data; they are merely indicators of trends.

Loudness as a Function of Frequency and Amplitude

The relationship between loudness and the frequency and SPL of the simplest of sounds, the pure tone, was first established by Fletcher and Munson, in 1933 [2]. There have been several subsequent redeterminations of loudness relationships by experimenters incorporating various refinements in their techniques. The data of Robinson and Dadson [3], for example, provide the basis for the International Organization for Standardization (ISO) recommendation R226 [4]. The presentation of loudness data is usually in the form of *equal-loudness contours*, as shown in Figure 1.4.2. Each curve shows the SPLs at which tones of various frequencies are judged to sound equal in loudness to a l-kHz reference tone; the SPL of the reference tone identifies the curve in units called *phons*. According to this method, the *loudness level* of a sound, in phons, is the SPL level of a l-kHz pure tone that is judged to be equally loud.

The equal-loudness contours of Figure 1.4.2 show that the ears are less sensitive to low frequencies than to middle and high frequencies and that this effect increases as sound level is reduced. In other words, as the overall sound level of a broadband signal such as music is reduced, the bass frequencies will fade faster than middle or high frequencies. In the curves, this appears as a crowding together of the contours at low frequencies, indicating that, at the lower sound levels, a small change in SPL of low-frequency sounds produces the same change in loudness as a larger change in SPL at middle and high frequencies. This may be recognized as the basis for the loudness compensation controls built into many hi-fi amplifiers, the purpose of

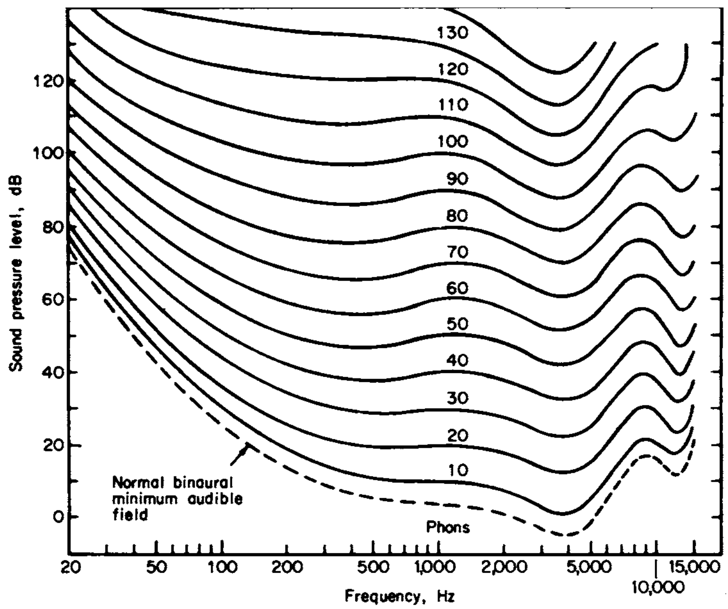


Figure 1.4.2 Contours of equal loudness showing the sound pressure level required for pure tones at different frequencies to sound as loud as a reference tone of 1000 Hz. (*From ISO Recommendation R226.*)

which is to boost progressively the bass frequencies as the overall sound level is reduced. The design and use of such compensation have often been erroneous because of a confusion between the shape of the loudness contours themselves and the *differences* between curves at various phon levels [5]. Sounds reproduced at close to realistic levels should need no compensation, since the ears will respond to the sound just as they would to the "live" version of the program. By the same token, control-room monitoring at *very* high sound levels can result in program equalization that is note appropriate to reproduction at normal domestic sound levels (combined with this are the effects of temporary and permanent changes in hearing performance caused by exposure to loud sounds).

It is difficult to take the interpretations of equal-loudness contours much beyond generalizations since, as mentioned earlier, they are composites of data from many individuals. There is also the fact that they deal with pure tones and the measurements were done either through headphones (Fletcher and Munson [2]) or in an anechoic chamber (Robinson and Dadson [3]). The relationship between these laboratory tests and the common application for these data, the audition of music in normal rooms, is one that is only poorly established.

The lowest equal-loudness contour defines the lower limit of perception: the *hearing-threshold level*. It is significant that the ears have their maximum sensitivity at frequencies that are important to the intelligibility of speech. This optimization of the hearing process can be seen in various other aspects of auditory performance as well.

The rate of growth of loudness as a function of the SPL is a matter of separate interest. Units of *sones* are used to describe the magnitude of the subjective sensation. One sone is defined as the loudness of a tone at the 40-phon loudness level. A sound of loudness 2 sones would be twice as loud, and a sound of 0.5 sone would be half as loud. The *loudness function* relating the subjective sensation to the physical magnitude has been studied extensively [6], and while there are consistencies in general behavior, there remain very large differences in the performance of individuals and in the effect of the temporal and spectral structure of the sound. A common approximation relates a change of 10 dB in SPL to a doubling or halving of loudness. Individual variations on this may be a factor of 2 or more, indicating that one is not dealing with precise data. For example, the growth of loudness at low frequencies, as shown in the curves of Figure 1.4.2, indicates a clear departure from the general rule. Nevertheless, it is worth noting that significant changes in loudness require large differences in SPL and sound power; a doubling of loudness that requires a l0-dB increase in sound level translates into a factor of 3.16 in sound pressure (or voltage) and a factor of 10 in power.

Loudness as a Function of Bandwidth

The studies of loudness that used pure tones leave doubts about how they relate to normal sounds that are complexes of several frequencies or continuous bands of sound extending over a range of frequencies. If the bandwidth of a sound is increased progressively while maintaining a constant overall measured sound level, it is found that loudness remains constant from narrow bandwidths up to a value called the *critical bandwidth*. At larger bandwidths, the loudness increases as a function of bandwidth because of a process known as *loudness summation*. For example, the broadband sound of an orchestra playing a chord will be louder than the simple sound of a flute playing a single note even when the sounds have been adjusted to the same SPL.

The critical bandwidth varies with the center frequency of the complex sound being judged. At frequencies below about 200 Hz it is fairly constant at about 90 Hz; at higher frequencies the critical bandwidth increases progressively to close to 4000 Hz at 15 kHz. The sound of the orchestra therefore occupies many critical bandwidths while the sound of the flute is predominantly within one band.

Loudness as a Function of Duration

Brief sounds can appear to be less loud than sounds with the same maximum sound level but longer duration. Experiments show that there is a progressive growth of loudness as signal duration is increased up to about 200 ms; above that, the relationship levels out. The implication is that the hearing system integrates sound energy over a time interval of about 200 ms. In reality, the integration is likely to be of neural energy rather than acoustical energy, which makes the process rather complicated, since it must embrace all the nonlinearities of the perceptual mechanism.

The practical consequence of this is that numerous temporal factors, such as duration, intermittency, repetition rate, and so on, all influence the loudness of sounds that are separate from SPL.

Measuring the Loudness of Complex Sounds

Given the numerous variables and uncertainties in ascertaining the loudness of simple sounds, it should come as no surprise that measuring the loudness of the wideband, complex, and ever-changing sounds of real life is a problem that has resisted simple interpretation. Motivated by the need to evaluate the annoyance value of sounds as well as the more neutral quantity of loudness,

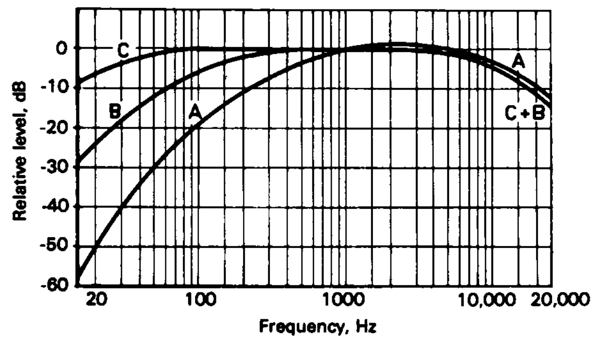


Figure 1.4.3 The standard frequency-weighting networks used in sound-level meters.

various methods have been developed for arriving at single-number ratings of complex sounds. Some methods make use of spectral analysis of the sound, adjusted by correction factors and weighting, to compute a single-number loudness rating. These tend to require expensive apparatus and are, at best, cumbersome to use; they also are most accurate for steady-state sounds.

Simplifying the loudness compensation permits the process to be accomplished with relatively straightforward electronics providing a direct-reading output in real time, a feature that makes the device practical for recording and broadcasting applications. Such devices are reported to give rather better indications of the loudness of typical music and speech program material than the very common and even simpler *volume-unit* (VU) meters or *sound-level meters* [7].

The VU meter responds to the full audio-frequency range, with a flat frequency response but with some control of its dynamic (time) response. A properly constructed VU meter should exhibit a response time of close to 300 ms, with an overswing of not more than 1.5 percent, and a return time similar to the response time. The dial calibrations and reference levels are also standardized. Such devices are therefore useful for measuring the magnitudes of steady-state signals and for giving a rough indication of the loudness of complex and time-varying signals, but they fail completely to take into account the frequency dependence of loudness.

The sound-level meters used for acoustical measurements are adjustable in both amplitude and time response. Various *frequency-weighting* curves, *A-weighting* being the most popular, acknowledge the frequency-dependent aspects of loudness, and “fast” and “slow” time responses deal differently with temporal considerations. Although these instruments are carefully standardized and find extensive use in acoustics, noise control, and hearing conservation, they are of limited use as program-level indicators. Figure 1.4.3 shows the common frequency-weighting options found in sound-level meters. A-weighting has become the almost universal choice for measurements associated with loudness, annoyance, and the assessment of hearing-damage risk.

Peak program meters (PPM) are also standardized [8], and they find extensive use in the recording and broadcast industries. However, they are used mainly as a means of avoiding overloading recorders and signal-processing equipment. Consequently, the PPM has a very rapid response (an integration time of about 10 ms in the normal mode), so that brief signal peaks are registered, and a slow return (around 3 s), so that the peak levels can be easily seen. These devices therefore are not useful indicators of loudness of fluctuating signals.

1.4.3b Masking

Listening to a sound in the presence of another sound, which for the sake of simplicity we shall call noise, results in the desired sound being, to some extent, less audible. This effect is called *masking*. If the noise is sufficiently loud, the signal can be completely masked, rendering it inaudible; at lower noise levels the signal will be partially masked, and only its apparent loudness may be reduced. If the desired sound is complex, it is possible for masking to affect only portions of the total sound. All this is dependent on the specific nature of both the signal and the masking sound.

In audio it is possible for the low-level sounds of music, for example, to be masked by background noise in a sound system. That same noise can mask distortion products, so the effects need not be entirely undesirable. In addition to the unwanted noises that have been implied so far, there can be masking of musical sounds by other musical sounds. Thus we encounter the interesting situation of the perceived sound of a single musical instrument modified by the sounds of other instruments when it is joined in an ensemble.

In addition to the partial and complete masking that occurs when two sounds occur simultaneously, there are instances of *temporal masking*, when the audibility of a sound is modified by a sound that precedes it in time (*forward masking*) or, strange as it may seem, by a sound that follows it (*backward masking*).

Simultaneous Masking

At the lowest level of audibility, the threshold, the presence of noise can cause a *threshold shift* wherein the amplitude of the signal must be increased to restore audibility. At higher sound levels the masked sound may remain audible but, owing to partial masking, its loudness can be reduced.

In *simultaneous masking* the signal and the masking sound coexist in the time domain. It is often assumed that they must also share the same frequency band. While this seems to be most effective, it is not absolutely necessary. The effect of a masking sound can extend to frequencies that are both higher and lower than those in the masking itself. At low sound levels a masking sound tends to influence signals with frequencies close to its own, but at higher sound levels the masking effect spreads to include frequencies well outside the spectrum of the masker. The dominant effect is an *upward spread* of masking that can extend several octaves above the frequency of the masking sound. There is also a *downward spread* of masking, but the effect is considerably less. In other words, a low-frequency masking sound can reduce the audibility of higher-frequency signals, but a high-frequency masking sound has relatively little effect on signals of lower frequency. Figure 1.4.4 shows that a simple masking sound elevates the hearing threshold over a wide frequency range but that the elevation is greater for frequencies above the masking sound.

In the context of audio, this means that we have built-in noise and distortion suppression. Background noises of all kinds are less audible while the music is playing but stand out clearly during the quiet intervals. Distortions generated in the recording and reproduction processes are present only during the musical sound and are therefore at least partially masked by the music itself. This is especially true for harmonic distortions, in which the objectionable distortion products are at frequencies higher than the masking sound—the sound that causes them to exist. Intermodulation-distortion products, on the other hand, are at frequencies both above and below

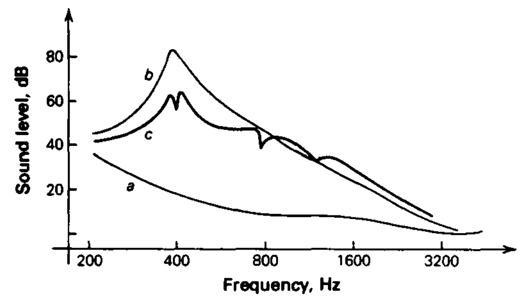


Figure 1.4.4 Detection threshold for pure tones of various frequencies: (*a*) in insolation, (*b*) in the presence of a narrow band (365 to 455 Hz) of masking noise centered on 400 Hz at a sound level of 80 dB, (*c*) in the presence of a making tone of 400 Hz at 80 dB. (*From* [40]. *Used with permission.*)

the frequencies of the signals that produce the distortion. In this case, the upper distortion products will be subject to greater masking by the signal than the lower distortion products.

Studies of distortion have consistently noted that all forms of distortion are less audible with music than with simple signals such as single tones or combinations of tones; the more effective masking of the spectrally complex music signal is clearly a factor in this. Also noted is that intermodulation distortion is more objectionable than its harmonic equivalent. A simple explanation for this may be that not only are the difference-frequency components of intermodulation distortion unmusical, but they are not well masked by the signals that produce them.

Temporal Masking

The masking that occurs between signals not occurring simultaneously is known as *temporal masking.* It can operate both ways, from an earlier to a later sound (forward masking) or from a later to an earlier sound (backward masking). The apparent impossibility of backward masking (going backward in time) has a physiological explanation. It takes time for sounds to be processed in the peripheral auditory system and for the neural information to travel to the brain. If the later sound is substantially more intense than the earlier sound, information about it can take precedence over information about the earlier sound. The effect can extend up to 100 to 200 ms, but because such occurrences are rare in normal hearing, the most noteworthy auditory experiences are related to forward masking.

Forward masking results from effects of a sound that remain after the physical stimulus has been removed. The masking increases with the sound level of the masker and diminishes rapidly with time, although effects can sometimes be seen for up to 500 ms [9]. Threshold shifts of 10 to 20 dB appear to be typical for moderate sound levels, but at high levels these may reach 40 to 50 dB. Combined with these substantial effects is a broadening of the frequency range of the masking; at masker sound levels above about 80 dB maximum masking no longer occurs at the frequency of the masking sound but at higher frequencies.

There are complex interactions among the numerous variables in the masking process, and it is difficult to translate the experimental findings into factors specifically related to audio engineering. The effects are not subtle, however, and it is clear that in many ways they influence what we hear.

1.4.3c Acoustic Reflex

One of the less-known features of hearing is the *acoustic reflex*, an involuntary activation of the middle-ear muscles in response to sound and some bodily functions. These tiny muscles alter the transmission of sound energy through the middle ear, changing the quantity and quality of the sound that reaches the inner ear. As the muscles tighten, there may be a slight reduction in the overall sound level reaching the inner ear, but mainly there is a change in spectral balance as the low frequencies are rolled off. Below approximately 1 kHz the attenuation is typically 5 to 10dB, but it can be as much as 30 dB.

The reflex is activated by sounds above 80- to 85-dB SPL, which led to the early notion that it was a protective mechanism; however, the most hazardous sounds are at frequencies that are little affected by the reflex, and, furthermore, the reflex is too slow to block the passage of loud transients. The reflex activates rather slowly, in 10 to 20 ms for loud sounds and up to 150 ms for sounds near the activation threshold; then, after an interval, it slowly relaxes. Obviously there have to be other reasons for its existence. Although there is still some speculation as to its purpose, the fact that it is automatically activated when we talk and when we chew suggests that part of the reason is simply to reduce the auditory effects of our own voice and eating sounds.

Some people can activate the reflex voluntarily, and they report a reduction in the loudness of low frequencies during the period of activation. The behavior of the reflex also appears to depend on the state of the listener's attention to the sound itself This built-in tone control clearly is a complication in sound quality assessments since the spectral balance appears to be a function of sound level, the pattern of sound-level fluctuations in time, and the listener's attitude or attention to the sound.

1.4.3d Pitch

Pitch is the subjective attribute of frequency, and while the basic correspondence between the two domains is obvious—low pitch to low frequencies and high pitch to high frequencies—the detailed relationships are anything but simple.

Fortunately waveforms that are periodic, however complex they may be, tend to be judged as having the same pitch as sine waves of the same repetition frequency. In other words, when a satisfactory pitch match has been made, the fundamental frequency of a complex periodic sound and a comparison sinusoid will normally be found to have the same frequency.

The exceptions to this simple rule derive from those situations where there is no physical energy at the frequency corresponding to the perceived pitch. Examples of pitch being associated with a *missing fundamental* are easily demonstrated by using groups of equally spaced tones, such as 100, 150, and 200 Hz, and observing that the perceived pitch corresponds to the difference frequency, 50 Hz. Common experience with sound reproducers, such as small radios, that have limited low-frequency bandwidth, illustrates the strength of the phenomenon, as do experiences with musical instruments, such as some low-frequency organ sounds, that may have little energy at the perceived fundamental frequency.

Scientifically, pitch has been studied on a continuous scale, in units of *mels*. It has been found that there is a highly nonlinear relationship between subjectively judged ratios of pitch and the corresponding ratios of frequency, with the subjective pitch interval increasing in size with increasing frequency. All this, though, is of little interest to traditional musicians, who have organized the frequency domain into intervals having special tonal relationships. The *octave* is particularly notable because of the subjective similarity of sounds spaced an octave apart and the fact that these sounds commonly differ in frequency by factors of 2. The musical fifth is a similarly well-defined relationship, being a ratio of 3:2 in repetition frequencies. These and the other intervals used in musical terminology gain meaning as one moves away from sine waves, with their one-frequency purity, into the sounds of musical instruments with their rich collection of overtones, many of which are harmonically related. With either pure tones [10] or some instrumental sounds in which not all the overtones are exactly harmonically related, the subjective octave may differ slightly from the physical octave; in the piano this leads to what is called *stretched tuning* [11].

The incompatibility of the mel scale of pitch and the hierarchy of musical intervals remains a matter for discussion. These appear to be quite different views of the same phenomenon, with some of the difference being associated with the musical expertise of listeners. It has been suggested, for example, that the mel scale might be better interpreted as a scale of *brightness* rather than one of pitch [12]. With periodic sounds brightness and pitch are closely related, but there are sounds, such as bells, hisses, and clicks, that do not have all the properties of periodic sounds and yet convey enough of a sense of pitch to enable tunes to be played with them, even though they cannot be heard as combining into chords or harmony. In these cases, the impressions of brightness and pitch seem to be associated with a prominence of sound energy in a band of frequencies rather than with any of the spectral components (partials, overtones, or harmonics) that may be present in the sound. A separate confirmation of this concept of brightness is found in subjective assessments of reproduced sound quality, where there appears to be a perceptual dimension along a continuum of “darkness” to “brightness” in which brightness is associated with a frequency response that rises toward the high frequencies or in which there are peaks in the treble [13]. At this, we reach a point in the discussion where it is more relevant to move into a different but related domain.

1.4.3e Timbre, Sound Quality, and Perceptual Dimensions

Sounds may be judged to have the same subjective dimensions of loudness and pitch and yet sound very different from one another. This difference in *sound quality*, known as *timbre* in musical terminology, can relate to the tonal quality of sounds from specific musical instruments as they are played in live performance, to the character of tone imparted to all sounds processed through a system of recording and reproduction, and to the tonal modifications added by the architectural space within which the original performance or a reproduction takes place. Timbre is, therefore, a matter of fundamental importance in audio, since it can be affected by almost anything that occurs in the production, processing, storage, and reproduction of sounds.

Timbre has many dimensions, not all of which have been confidently identified and few of which have been related with any certainty to the corresponding physical attributes of sound. There is, for example, no doubt that the shape and composition of the frequency spectrum of the sound are major factors, as are the temporal behaviors of individual elements comprising the spectrum, but progress has been slow in identifying those measurable aspects of the signal that correlate with specific perceived dimensions, mainly because there are so many interactions between the dimensions themselves and between the physical and psychological factors underlying them.

The field of electronic sound synthesis has contributed much to the understanding of why certain musical instruments sound the way they do, and from this understanding have followed devices that permit continuous variations of many of the sound parameters. The result has been progressively better imitations of acoustical instruments in electronic simulations, as well as an infinite array of new "instruments" exhibiting tonal colors, dynamics, and emotional connotations that are beyond the capability of traditional instruments. At the same time as this expansion of timbral variety is occurring on one front of technical progress, there is an effort on another front to faithfully preserve the timbre of real and synthesized instruments through the complex process of recording and reproduction.

A fundamental problem in coming to grips with the relationship between the technical descriptions of sounds and the perception of timbre is in establishing some order in the choice and quantitative evaluation of words and phrases used by listeners to describe aspects of sound quality. Some of the descriptors are fairly general in their application and seem to fall naturally to quantification on a continuous scale from say, "dull" to "bright" or from "full" to "thin." Others, though, are specific to particular instruments or lapse into poetic portrayals of the evoked emotions.

From carefully conducted assessments of reproduced sound quality involving forms of multivariate statistical analysis, it has become clear that the extensive list can be reduced to a few relatively independent dimensions. As might be expected, many of the descriptors are simply different ways of saying the same thing, or they are responses to different perceptual manifestations of the same physical phenomenon.

From such analyses can come useful clarifications of apparently anomalous results since these responses need not be unidirectional. For example, a relatively innocent rise in the high-frequency response of a sound reproducer might be perceived as causing violins to sound unpleasantly strident but cymbals to sound unusually clear and articulate. A nice sense of air and space might be somewhat offset by an accentuation of background hiss and vocal sibilants, and so on.

Inexperienced listeners tend to concentrate unduly on a few of the many descriptors that come to mind while listening, while slightly more sophisticated subjects may become confused by the numerous contradictory indications. Both groups, for different reasons, may fail to note that there is but a single underlying technical flaw. The task of critical listening is one that requires a broad perspective and an understanding of the meaning and relative importance of the many timbral clues that a varied musical program can reveal. Trained and experienced listeners tend to combine timbral clues in a quest for logical technical explanations for the perceived effects. However, with proper experimental controls and the necessary prompting through carefully prepared instructions and a questionnaire, listeners with little prior experience can arrive at similar evaluations of accuracy without understanding the technical explanations [14].

The following list of perceptual dimensions is derived from the work of Gabrielsson and various colleagues [13, 15], and is the basis for listening questionnaires used extensively by those workers and the author [14]. The descriptions are slightly modified from the original [13].

- *Clarity, or definition*: This dimension is characterized by adjectives such as clear, well defined, distinct, clean or pure, and rich in details or detailed, as opposed to adjectives such as diffuse, muddy or confused, unclear, blurred, noisy, rough, harsh, or sometimes rumbling, dull, and faint. High ratings in this dimension seem to require that the reproduction system perform well in several respects, exhibiting a wide frequency range, flat frequency response, and low nonlinear distortion. Systems with limited bandwidth, spectral irregularities due to resonances, or audible distortion receive lower ratings. Low-frequency spectral emphasis seems also to be detrimental to performance in this dimension, resulting in descriptions of rumbling, for the obvious reason, and dullness, probably due to the upward masking effects of the strong low frequencies. Increased sound levels result in increased clarity and definition.

- *Sharpness, or hardness, versus softness*: Adjectives such as sharp, hard, shrill, screaming, pointed, and clashing are associated with this dimension, contrasted with the opposite qualities of soft, mild, calm or quiet, dull, and subdued. A rising high-frequency response or prominent resonances in the high-frequency region can elicit high ratings in this dimension, as can certain forms of distortion. A higher or lower sound level also contributes to movement within this dimension, with reduced levels enhancing the aspect of softness.
- *Brightness versus darkness*: This dimension is characterized by the adjective bright, as opposed to dark, rumbling, dull, and emphasized bass. There appears to be a similar relationship between this dimension and the physical attributes of the sound system as exists with the preceding dimension, sharpness, or hardness, versus softness. In experiments, the two dimensions sometimes appear together and sometimes separately. The sense of pitch associated with brightness might be a factor in distinguishing between these two dimensions.
- *Fullness versus thinness*: This dimension also can appear in combination with brightness versus darkness, and there are again certain similarities in the relationship to measured spectrum balance and smoothness. There appears to be an association with the bandwidth of the system, especially at the low frequencies, and with sound level. It seems possible that this dimension is a representation of one encountered elsewhere as volume, which has been found to increase with increasing sound level but to decrease with increasing frequency.
- *Spaciousness*: Almost self-explanatory, this dimension elicits expressions of spacious, airy, wide, and open, as opposed to closed or shut up, narrow, and dry. The phenomenon appears to be related to poorly correlated sounds at the two ears of the listener. Other aspects of spaciousness are related to the spectrum of the reproduced sound. Gabrielsson points out that increased treble response enhances spaciousness, while reducing the bandwidth encourages a closed or shut-up impression. It is well known that the directional properties of the external ear (Figure 1.4.5) encode incoming sounds with spectral cues that can be significant influences in sound localization [16]. One such cue is a moving spectral notch and an increase in the sound level reaching the eardrum over the band from 5 to 10 kHz for progressively elevated sources (Figure 1.4.6). The appropriate manipulation of the sound spectrum in this frequency region can alone create impressions of height [17, 18] and, in this sense, alter the impression of spaciousness. It is worthy of note that the dimension of spaciousness is clearly observed in monophonic as well as stereophonic reproductions, indicating that it is a rather fundamental aspect of sound quality [13, 19].
- *Nearness*: Differences in the apparent proximity of sound sources are regularly observed in listening tests. It is clear that sound level affects perception of distance, especially for sounds such as the human voice that are familiar to listeners. Evidence from other studies indicates that impressions of distance are also influenced by the relationship between the direct, early-reflected, and reverberant sounds and the degree of coherence that exists in these sounds as they appear at the listener's ears [17].
- *Absence of extraneous sounds*: This dimension refers to nonmusical sounds that either exist in the original program material and are accentuated by aspects of the reproducer (such as tape hiss being aggravated by a treble boost) or are generated within the device itself (such as electronic amplifier clipping or mechanical noises from a loudspeaker).

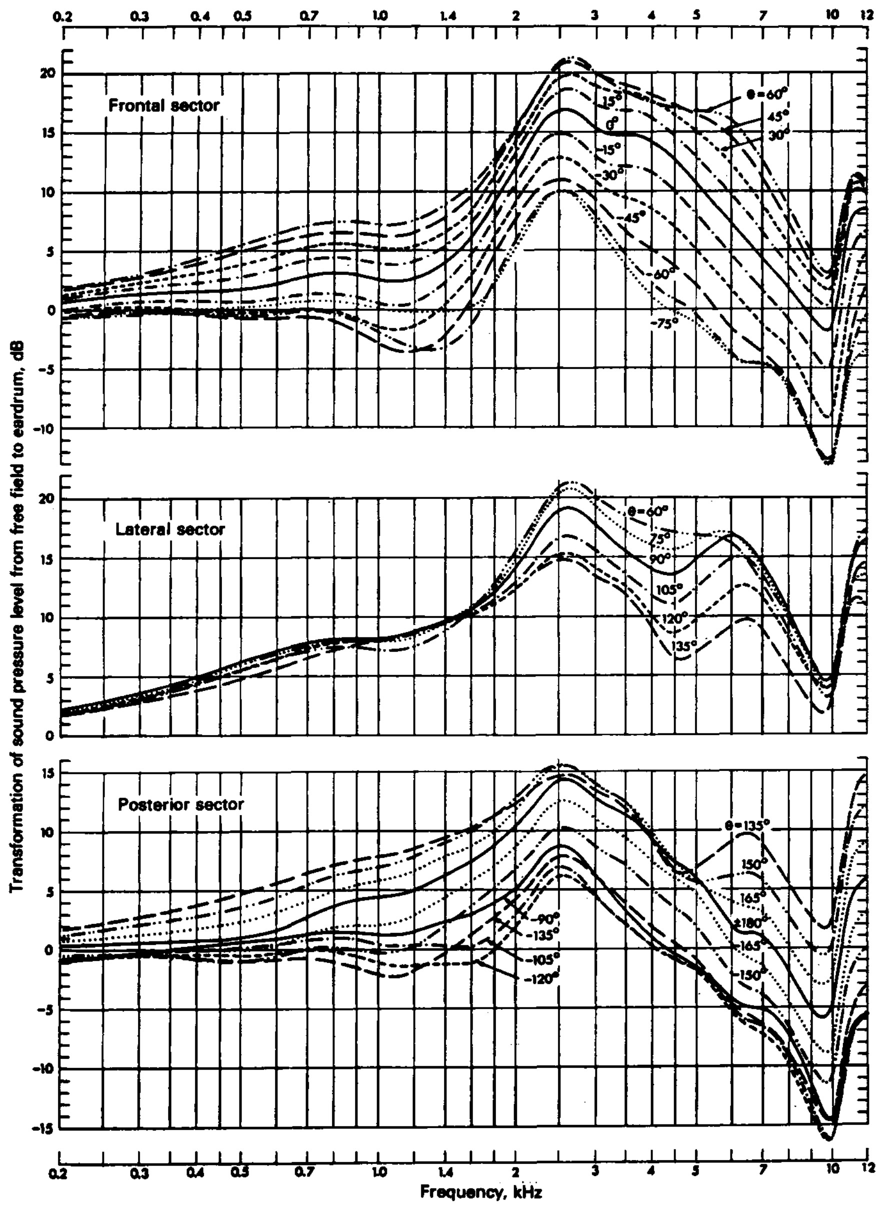


Figure 1.4.5 Family of curves showing the transformation of sound pressure level from the free field to the eardrum in the horizontal plane as a function of frequency, averaged over many listeners in several independent studies. The horizontal angles are referred to zero (the forward direction) and increase positively toward the ear in which the measurement is made and negatively away from it. (*From* [27]. *Used with permission.*)

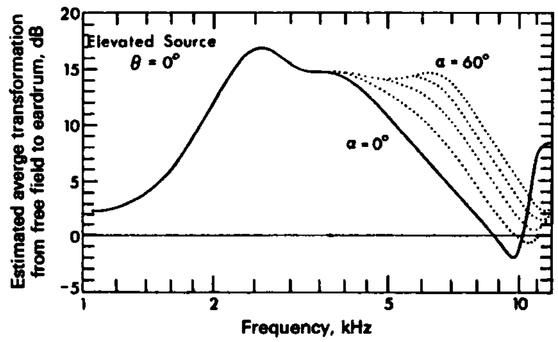


Figure 1.4.6 The estimated average transformation of sound pressure level from the free field to the eardrum as a function of frequency, showing the variations as a function of the angle of elevation for sounds arriving from the forward direction. (*From* [1]. *Used with permission.*)

- *Loudness*: This self-explanatory dimension is a useful check on the accuracy with which the sound levels of comparison sounds have been matched. It should, however, be noted that some listeners seem to regard the adjective loud as a synonym for sharp, hard, or painful.

The relative importance of these dimensions in describing overall sound quality changes slightly according to the specific nature of the devices under test, the form of the listener questionnaire, the program material, and, to some extent, the listeners themselves. In general, Gabrielsson and colleagues [13, 15] have found that clarity, or definition, brightness versus darkness, and sharpness, or hardness, versus softness are major contributors to the overall impression of sound quality.

1.4.3f Audibility of Variations in Amplitude and Phase

Other things being equal, very small differences in sound level can be heard: down to a fraction of a decibel in direct A/B comparisons. Level differences that exist over only a small part of the spectrum tend to be less audible than differences that occupy a greater bandwidth. In other words, a small difference that extends over several octaves may be as significant as a much larger difference that is localized in a narrow band of frequencies. Spectral tilts of as little as 0.1 dB per octave are audible. For simple sounds the only audible difference may be loudness, but for complex sounds differences in timbre may be more easily detectable.

The audibility of phase shift is a very different matter. Several independent investigations over many years have led to the conclusion that while there are some special signals and listening situations where phase effects can be heard, their importance when listening to music in conventional environments is small [19]. Psychophysical studies indicate that, in general, sensitivity to phase is small compared with sensitivity to the amplitude spectrum and that sensitivity to phase decreases as the fundamental frequency of the signal increases. At the same time, it appears to be phase shifts in the upper harmonics of a complex signal that contribute most to changes in timbre [20].

The notion that phase, and therefore waveform, information is relatively unimportant is consistent with some observations of normal hearing. Sounds from real sources (voices and musical instruments) generally arrive at our ears after traveling over many different paths, some of which may involve several reflections. The waveform at the ear therefore depends on various factors other than the source itself. Even the argument that the direct sound is especially selected for audition and that later arrivals are perceptually suppressed does not substantially change the situation because sources themselves do not radiate waveforms that are invariably distinctive. With musical instruments radiating quite different components of their sound in different directions (consider the complexity of a grand piano or the cello, for example), the sum of these components—the waveform at issue—will itself be different at every different angle and distance; a recording microphone is in just such a situation.

The fact that the ear seems to be relatively insensitive to phase shifts would therefore appear to be simply a condition born of necessity. It would be incorrect to assume, however, that the phase performance of devices is *totally* unimportant. Spectrally localized phase anomalies are useful indicators of the presence of resonances in systems, and very large accumulations of phase shift over a range of frequencies can become audible as group delays.

While the presence of resonances can be inferred from phase fluctuations, their audibility may be better predicted from evidence in the amplitude domain [19]. It should be added that resonances of low *Q* in sound reproduction systems are more easily heard than those of higher Q [21–23]. This has the additional interesting ramification that evidence of sustained ringing in the time domain may be less significant than ringing that is rapidly damped; waveform features and other measured evidence that attract visual attention do not always correspond directly with the sound colorations that are audible in typical listening situations.

1.4.3g Perception of Direction and Space

Sounds are commonly perceived as arriving from specific directions, usually coinciding with the physical location of the sound source. This perception may also carry with it a strong impression of the acoustical setting of the sound event, which normally is related to the dimensions, locations, and sound-reflecting properties of the structures surrounding the listener and the sound source as well as objects in the intervening path.

Blauert, in his thorough review of this field [17], defines *spatial hearing* as embracing "the relationships between the locations of auditory events and other parameters—particularly those of sound events, but also others such as those that are related to the physiology of the brain." This statement introduces terms and concepts that may require some explanation. The adjective *sound*, as in *sound event*, refers to a physical source of sound, while the adjective *auditory* identifies a perception. Thus, the perceived location of an auditory event usually coincides with the physical location of the source of sound. Under certain circumstances, however, the two locations may differ slightly or even substantially. The difference is then attributed to other parameters having nothing whatever to do with the physical direction of the sound waves impinging on the ears of the listener, such as subtle aspects of a complex sound event or the processing of the sound signals within the brain.

Thus have developed the parallel studies of *monaural*, or one-eared, hearing and *binaural*, or two-eared, hearing. Commercial sound reproduction has stimulated a corresponding interest in the auditory events associated with sounds emanating from a single source (*monophonic*) and from multiple sources that may be caused to differ in various ways (*stereophonic*). In common usage it is assumed that stereophonic reproduction involves only two loudspeakers, but there are many other possible configurations. In stereophonic reproduction the objective is to create many more auditory events than the number of real sound sources would seem to permit. This is accomplished by presenting to the listener combinations of sounds that take advantage of certain inbuilt perceptual processes in the brain to create auditory events in locations other than those of the sound events and in auditory spaces that may differ from the space within which the reproduction occurs.

Understanding the processes that create auditory events would ideally permit the construction of predictable auditory spatial illusions in domestic stereophonic reproduction, in cinemas, in concert halls, and in auditoria. Although this ideal is far from being completely realized, there are some important patterns of auditory behavior that can be used as guides for the processing of sound signals reproduced through loudspeakers as well as for certain aspects of listening room, concert hall, and auditorium design.

1.4.3h Monaural Transfer Functions of the Ear

Sounds arriving at the ears of the listener are subject to modification by sound reflection, diffraction, and resonances in the structures of the external ear, head, shoulders, and torso. The amount and form of the modification are dependent on the frequency of the sound and the direction and distance of the source from which the sound emanates. In addition to the effect that this has on the sensitivity of the hearing process, which affects signal detection, there are modifications that amount to a kind of directional encoding, wherein sounds arriving from specific directions are subject to changes characteristic of those directions.

Each ear is partially sheltered from sounds arriving from the other side of the head. The effect of diffraction is such that low-frequency sounds, with wavelengths that are large compared with the dimensions of the head, pass around the head with little or no attenuation, while higher frequencies are progressively more greatly affected by the directional effects of diffraction. There is, in addition, the acoustical interference that occurs among the components of sound that have traveled over paths of slightly different length around the front and back and over the top of the head.

Superimposed on these effects are those of the pinna, or external ear. The intriguingly complex shape of this structure has prompted a number of theories of its behavior, but only relatively recently have some of its important functions been properly put into perspective. According to one view, the folds of the pinna form reflecting surfaces, the effect of which is to create, at the entrance to the ear canal, a system of interferences between the direct and these locally reflected sounds that depends on the direction and distance of the incoming sound [24]. The small size of the structures involved compared with the wavelengths of audible sounds indicates that dispersive scattering, rather than simple reflection, is likely to be the dominant effect. Nevertheless, measurements have identified some acoustical interferences resembling those that such a view would predict, and these have been found to correlate with some aspects of localization [18, 25].

In the end, however, the utility of the theory must be judged on the basis of how effectively it explains the physical functions of the device and how well it predicts the perceptual consequences of the process. From this point of view, time-domain descriptions would appear to be at a disadvantage since the hearing process is demonstrably insensitive to the fine structure of signals at frequencies above about 1.5 kHz [17]. Partly for this reason most workers have favored descriptions in terms of spectral cues.

It is therefore convenient that the most nearly complete picture of external-ear function has resulted from examinations of the behavior of the external ear in the frequency domain. By carefully measuring the pressure distributions in the standing-wave patterns, the dominant resonances in the external ear have been identified [26.] These have been related to the physical structures and to the measured acoustical performance of the external ear [1].

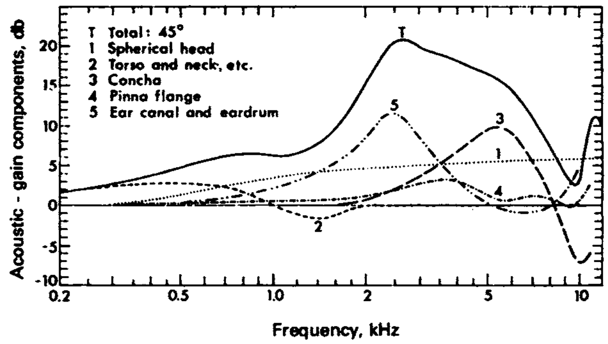


Figure 1.4.7 Contributions of various body parts to the total acoustic gain of the external hearing system for a sound source at a horizontal angle of 45°. Note that the interactions between these components prevent simple arithmetic addition of their individual contributions. (*From* [1]. *Used with permission.*)

A particularly informative view of the factors involved in this discussion comes from an examination of curves showing the transformation of SPL from the free field to the eardrum [27]. These curves reveal, as a function of frequency, the amplitude modifications imposed on incident sounds by the external hearing apparatus. Figure 1.4.5 shows the family of curves representing this transformation for sounds arriving from different directions in the horizontal plane. Figure 1.4.6 shows the estimated transformations for sound sources at different elevations.

An interesting perspective on these data is shown in Figure 1.4.7, where it is possible to see the contributions of the various acoustical elements to the total acoustical gain of the ear. It should be emphasized that there is substantial acoustical interaction among these components, so that the sum of any combination of them is not a simple arithmetic addition. Nevertheless, this presentation is a useful means of acquiring a feel for the importance of the various components.

It is clear from these curves that there are substantial direction-dependent spectral changes, some rather narrowband in influence and others amounting to significant broadband tilts. Several studies in localization have found that, especially with pure tones and narrowband signals, listeners could attribute direction to auditory events resulting from sounds presented through only one ear (monaural localization) or presented identically in two ears, resulting in localization in the *median plane* (the plane bisecting the head vertically into symmetrical left-right halves). So strong are some of these effects that they can cause auditory events to appear in places different from the sound event, depending only on the spectral content of the sound. Fortunately such confusing effects are not common in the panorama of sounds we normally encounter, partly because of familiarity with the sounds themselves, but the process is almost certainly a part of the mechanism by which we are able to distinguish between front and back and between up an down, directions that otherwise would be ambiguous because of the symmetrical locations of the two ears.

Interaural Differences

As useful as the monaural cues are, it is sound localization in the horizontal plane that is dominant, and for this the major cues come from the comparison of the sounds at the two ears and the

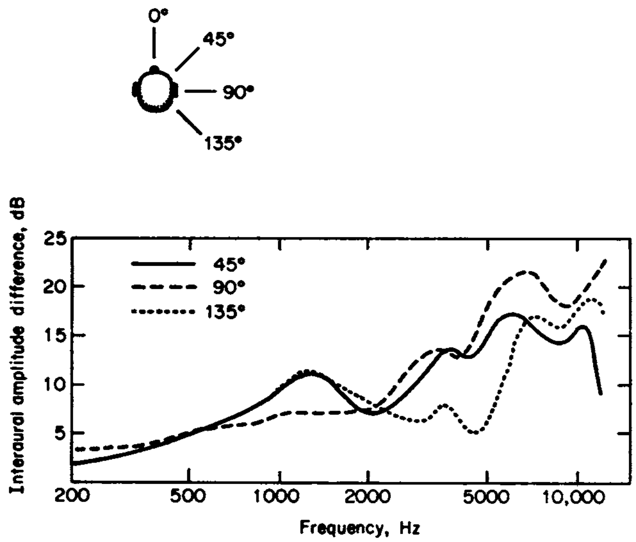


Figure 1.4.8 The interaural amplitude difference as a function of frequency for three angles of incidence. (*After* [28].)

analysis of the differences between them. From the data shown in Figure 1.4.5 it is evident that there is a substantial frequency-dependent *interaural amplitude difference* (lAD) that characterizes sounds arriving from different horizontal angles. Because of the path length differences there will also be an associated *interaural time difference* (ITD) that is similarly dependent on horizontal angle.

Figure 1.4.8 shows IADs as a function of frequency for three angles of incidence in the horizontal plane. These have been derived from the numerical data in [28], from which many other such curves can be calculated.

The variations in IAD as a function of both frequency and horizontal angle are natural consequences of the complex acoustical processes in the external hearing apparatus. Less obvious is the fact that there is frequency dependency in the ITDs. Figure 1.4.9 shows the relationship between ITD and horizontal angle for various pure tones and for broadband clicks. Also shown are the predictive curves for low-frequency sounds, based on diffraction theory, and for high-frequency sounds, based on the assumption that the sound reaches the more remote ear by traveling as a creeping wave that follows the contour of the head. At intermediate frequencies (0.5 to 2 kHz) the system is dispersive, and the temporal differences become very much dependent on the specific nature of the signal [29, 30].

It is evident from these data that at different frequencies, especially the higher frequencies, there are different combinations of ITD and IAD associated with each horizontal angle of incidence. Attempts at artificially manipulating the localization of auditory events by means of frequency-independent variations of these parameters are therefore unlikely to achieve the image size and positional precision associated with natural sound events.

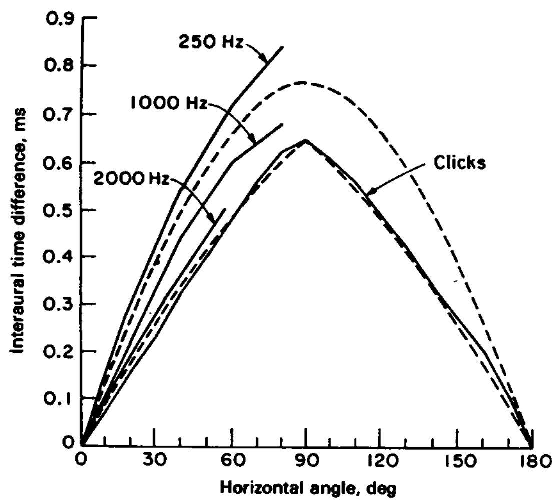


Figure 1.4.9 Interaural time difference as a function of horizontal angle. The curves show measured data for clicks and pure tones (solid lines) and predictive curves for low frequencies (top dashed curve), based on diffraction theory, and for high frequencies (bottom dashed curve), based on creeping-wave concepts. (*From* [41]. *Used with permission.*)

Localization Blur

In normal hearing the precision with which we are able to identify the direction of sounds depends on a number of factors. The measure of this precision is called *localization blur*, the smallest displacement of the sound event that produces a just-noticeable difference in the corresponding auditory event. The concept of localization blur characterizes the fact that auditory space (the perception) is less precisely resolved than physical space and the measures we have of it.

The most precise localization is in the horizontal forward direction with broadband sounds preferably having some impulsive content. The lower limit of localization blur appears to be about 1°, with typical values ranging from 1 to 3°, though for some types of sound values of 10° or more are possible. Moving away from the forward axis, localization blur increases, with typical values for sources on either side of the head and to the rear being around 10 to 20°. Vertically, localization blur is generally rather large, ranging from about 5 to 20° in the forward direction to 30 to 40° behind and overhead [17].

Lateralization versus Localization

In exploring the various ways listeners react to interaural signal differences, it is natural that headphones be used, since the sounds presented to the two ears can then be independently controlled. The auditory events that result from this process are distinctive, however, in that the perceived images occur inside or very close to the head and image movement is predominantly lateral. Hence, this phenomenon has come to be known as *lateralization*, as opposed to *localization*, which refers to auditory events perceived to be external and at a distance. Overcoming the

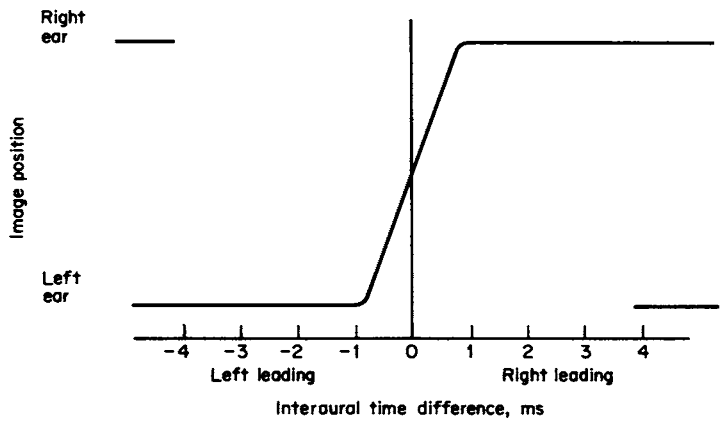


Figure 1.4.10 Perceived positions of the dominant auditory images resulting from impulsive signals (clicks) presented through headphones when the interaural time difference is varied.

in-head localization characteristic of headphone listening has been a major difficulty, inhibiting the widespread use of these devices for critical listening.

In headphone listening it is possible to move the auditory event by independently varying the interaural time or amplitude difference. Manipulating interaural time alone yields auditory image trajectories of the kind shown in Figure 1.4.10, indicating that the ITD required to displace the auditory image from center completely to one side is about 0.6 ms, a value that coincides with the maximum ITD occurring in natural hearing (Figure 1.4.9). Although most listeners would normally be aware of a single dominant auditory image even when the ITD exceeds this normal maximum value, it is possible for there to be multiple auditory images of lesser magnitude, each with a distinctive tonal character and each occupying a different position in perceptual space. With complex periodic signals, experienced listeners indicate that some of these images follow trajectories appropriate to the individual harmonics for frequencies that are below about 1 kHz [31]. This spatial complexity would not be expected in normal listening to a simple sound source, except when there are delayed versions of the direct sounds caused by strong reflections or introduced electronically. The result, if there are several such delayed-sound components, is a confused and spatially dispersed array of images, coming and going with the changing spectral and temporal structure of the sound. It seems probable that this is the origin of the often highly desirable sense of spaciousness in live and reproduced musical performances.

The sensitivity of the auditory system to changes in ITD in the lateralization of auditory images, or *lateralization blur* is dependent on both the frequency and the amplitude of the signal. According to various experimenters, lateralization blur varies from around 2 μs to about 60 μs, increasing as a function of signal frequency and sound level, and is at a minimum point around ITD = 0.

Introducing an IAD displaces the auditory event toward the ear receiving the louder sound. An IAD of between 10 and 20 dB seems to be sufficient to cause the image to be moved completely to one side. The precise figure is difficult to ascertain because of the rapid increase in lateralization blur as a function of IAD; the auditory event becomes wider as it approaches the side of the head. Close to center, however, the lateralization blur is consistently in the vicinity of 1 to 2 dB.

Spatial Impression

Accompanying the auditory impression of images in any normal environment is a clear impression of the type and size of the listening environment itself. Two aspects appear to be distinguishable: *reverberance*, associated with the temporal stretching and blurring of auditory events caused by reverberation and late reflections; and *spaciousness*, often described as a spreading of auditory events so that they occupy more space than the physical ensemble of sound sources. Other descriptors such as ambience, width, or envelopment also apply. Spaciousness is a major determinant of listener preference in concert halls and as such has been the subject of considerable study.

In general, the impression of spaciousness is closely related to a lack of correlation between the input signals to the two ears. This appears to be most effectively generated by strong early lateral reflections (those arriving within about the first 80 ms after the direct sound). While all spectral components appear to add positively to the effect and to listener preference, they can contribute differently. Frequencies below about 3 kHz seem to contribute mainly to a sense of depth and envelopment, while high frequencies contribute to a broadening of the auditory event [32].

The acoustical interaction of several time-delayed and directionally displaced sounds at the ears results in a reduced interaural cross correlation; the sense of spaciousness is inversely proportional to this correlation. In other terms, there is a spectral and temporal incoherence in the sounds at the ears, leading to the fragmentation of auditory events as a function of both frequency and time. The fragments are dispersed throughout the perceptual space, contributing to the impression of a spatially extended auditory event.

1.4.3i Distance Hearing

To identify the distance of a sound source listeners appear to rely on a variety of cues, depending on the nature of the sound and the environment. In the absence of strong reflections, as a sound source is moved farther from a listener, the sound level diminishes. It is possible to make judgments of distance on this factor alone, but only for sounds that are familiar, where there is a memory of absolute sound levels to use as a reference. With any sound, however, this cue provides a good sense of relative distance.

In an enclosed space the listener has more information to work with, because as a sound source is moved away, there will be a change in the relationship between the direct sound and the reflected and reverberant sounds in the room. The hearing mechanism appears to take note of the relative strengths of the direct and indirect sounds in establishing the distance of the auditory event. When the sound source is close, the direct sound is dominant and the auditory image is very compact; at greater distances, the indirect sounds grow proportionately stronger until eventually they dominate. The size of the auditory event increases with distance, as does the localization blur.

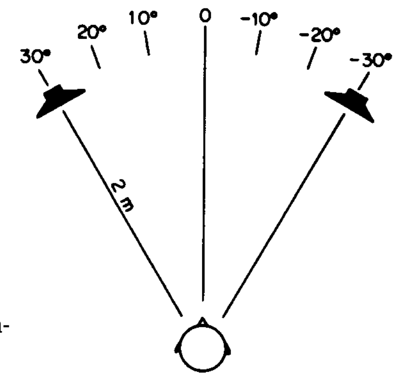


Figure 1.4.11 Standard stereophonic listening configuration.

1.4.3j Stereophonic Imaging

Consider the conventional stereophonic arrangement shown in Figure 1.4.11. If the two loudspeakers are radiating coherent sounds with identical levels and timing, the listener should perceive a single auditory event midway between the loudspeakers. This phantom, or virtual, sound source is the result of *summing localization*, the basis for the present system of two-channel stereophonic recording and reproduction.

Progressively increasing the time difference between the signals in the channels displaces the auditory event, or image, toward the side radiating the earlier sound until, at about 1 ms, the auditory image is coincident with the source of the earlier sound. At time differences greater than about 1 ms the perception may become spatially more dispersed, but the principal auditory event is generally perceived to remain at the position of the earlier sound event until, above some rather larger time difference, there will be two auditory events occurring separately in both time and space, the later of which is called an *echo*.

The region of time difference between that within which simple summing localization occurs and that above which echoes are perceived is one of considerable interest and complexity. In this region the position of the dominant auditory event is usually determined by the sound source that radiates the first sound to arrive at the listener's location. However, depending on the nature of the signal, simple summing localization can break down and there can be subsidiary auditory images at other locations as well. The later sound arrivals also influence loudness, timbre, and intelligibility in ways that are not always obvious.

The cause of this complexity can be seen in Figure 1.4.12, showing the sounds arriving at the two ears when the sound is symbolically represented by a brief impulse. It is immediately clear that the fundamental difference between the situation of summing localization and that of natural localization is the presence of four sound components at the ears instead of just two.

In all cases the listener responds to identical ear input signals by indicating a single auditory event in the forward direction. Note, however, that in both stereo situations the signals at the two ears are not the same as the signals in normal localization. Thus, even though the spatial aspects have been simulated in stereo, the sounds at the two ears are modified by the *acoustical crosstalk* from each speaker to the opposite ear, meaning that perfectly accurate timbral reproduction for these sounds is not possible. This aspect of stereo persists through all conditions for time-difference manipulation of the auditory image, but with amplitude-difference manipulation the effect

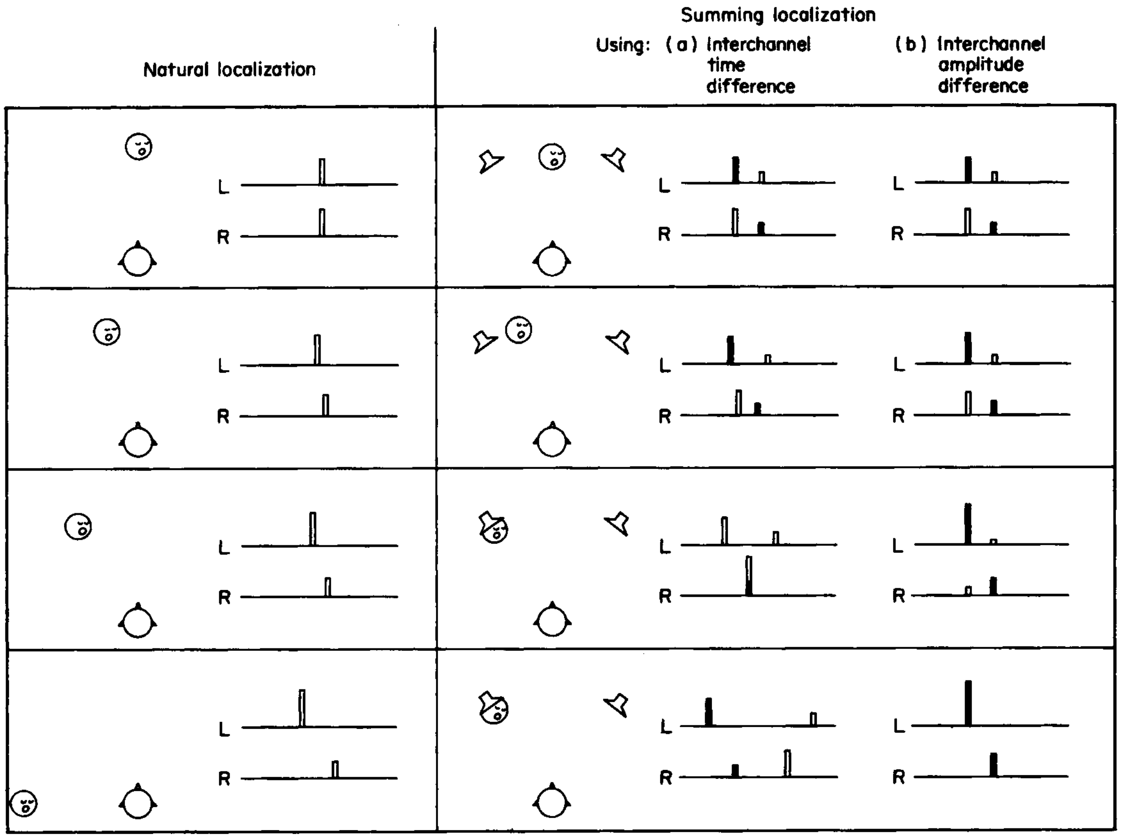


Figure 1.4.12 Comparison between sound localization in natural listening and localization in stereophonic listening within the range of simple summing. For the purposes of this simplified illustration, the sound waveform is an impulse. To the right of the pictoral diagrams showing a listener receiving sound from either a single source (natural localization) or a stereo pair of loudspeakers (summing localization) are shown the sounds received by the left and right ears of the listener. In the stereo illustrations, sounds from the left loudspeaker are indicated by dark bars and sounds from the right loudspeaker by light bars.

diminishes with increasing amplitude difference until, in the extreme, the listener hears only sound from a single speaker, a monophonic presentation.

Although impressions of image movement between the loudspeakers can be convincingly demonstrated by using either interchannel time or amplitude differences, there is an inherent limitation in the amount of movement: in both cases the lateral displacement of the principal auditory event is bounded by the loudspeakers themselves.

With time differences, temporal masking inhibits the contributions of the later arrivals, and the localization is dominated by the first sound to arrive at each ear. With small time differences the image can be moved between the loudspeakers, the first arrivals are from different loudspeakers, and it can be seen that an interchannel time difference is perceived as an ITD. At larger values of interchannel time difference the first arrivals are from the same loudspeaker, and the dominant auditory image remains at that location. This is because of the *law of the first wave-front*, also known as the precedence effect, according to which the dominant auditory event is perceived to be coincident with the loudspeaker radiating the earlier sound. The other sound components are still there nonetheless, and they can contribute to complexity in the spatial illusion as well as to changes in timbre.

With amplitude differences (also known as *intensity stereo*), the temporal pattern of events in the two ears is unchanged until the difference approaches infinity. At this point, the ears receive signals appropriate to a simple sound source with the attendant sound and localization accuracy. It is a real (monophonic) sound source generating a correspondingly real auditory event.

1.4.3k Summing Localization with Interchannel Time and Amplitude Differences

Figure 1.4.13 shows the position of the auditory image as a function of interchannel time difference for the conventional stereophonic listening situation shown in Figure 1.4.11. The curves shown are but a few of the many that are possible since, as is apparent, the trajectory of the auditory image is strongly influenced by signal type and spectral composition.

In contrast, the curves in Figure 1.4.14, showing the position of the auditory image as a function of interchannel amplitude difference, are somewhat more orderly. Even so, there are significant differences in the slopes of the curves for different signals.

With a signal like music that is complex in all respects, it is to be expected that, at a fixed time or amplitude difference, the auditory event will not always be spatially well defined or positionally stable. There are situations where experienced listeners can sometimes identify and independently localize several coexisting auditory images. Generally, however, listeners are inclined to respond with a single compromise localization, representing either the "center of gravity" of a spatially complex image display or the dominant component of the array. If the spatial display is ambiguous, there can be a strong flywheel effect in which occasional clear spatial indications from specific components of the sound engender the perception that all of that sound is continuously originating from a specific region of space. This is especially noticeable with the onset of transient or any small mechanical sounds that are easily localized compared with the sustained portion of the sounds.

The blur in stereo localization, as in natural localizaton, is least for an image localized in the forward direction, where, depending on the type of sound, the *stereo localization blur* is typically about 3 to 7°. With the image fully displaced by amplitude difference (IAD = 30 dB), the blur increases to typical values of 5 to 11°. With the image fully displaced by means of time difference (ITD = 1 ms), the blur increases to typical values of 10 to 16° [17].

Effect of Listener Position

Sitting away from the line of symmetry between the speakers causes the central auditory images to be displaced toward the nearer loudspeaker. Interaural time differences between the sound arrivals at the ears are introduced as the path lengths from the two speakers change. Within the first several inches of movement away from the axis of symmetry, the sound components from the left and right loudspeakers remain the first arrivals at the respective ears. In this narrow region it is possible to compensate for the effect of the ITD by adding the appropriate opposite bias of interchannel amplitude difference (see Figure 1.4.15). This process is known as *time-*

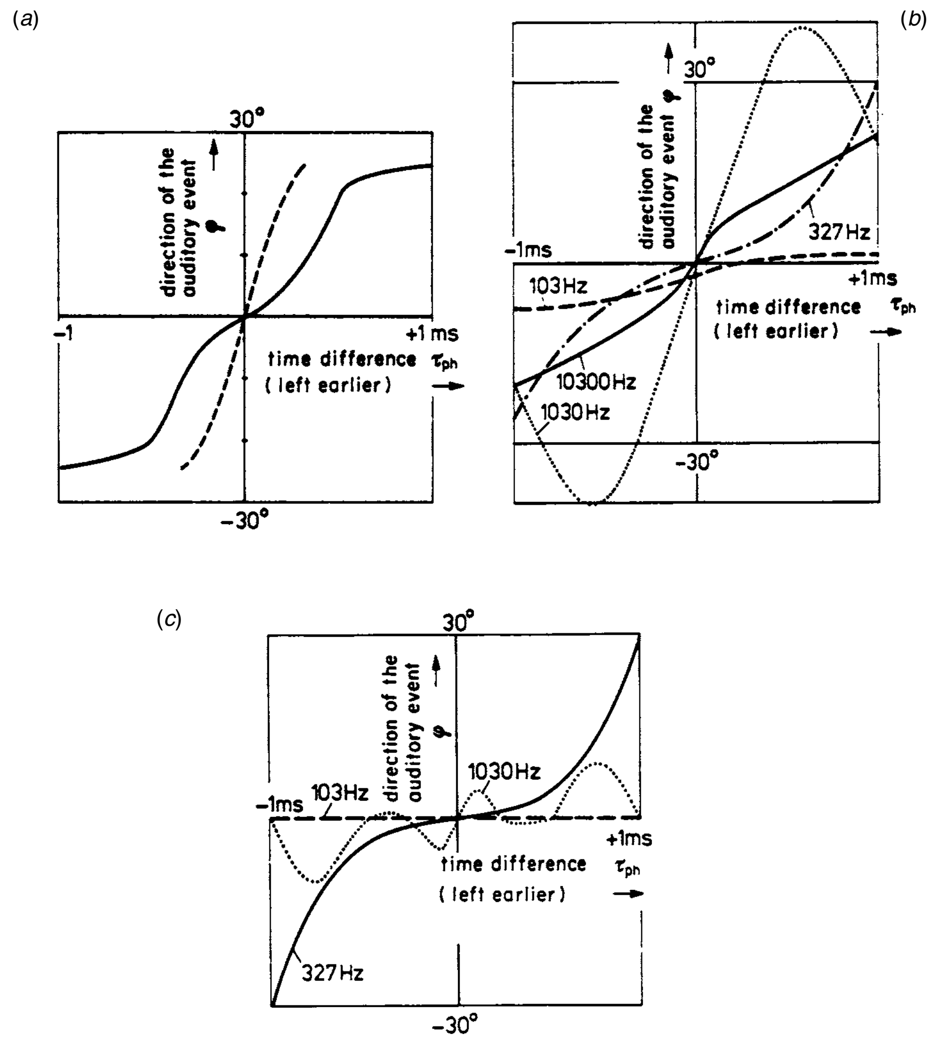


Figure 1.4.13 Direction of auditory images perceived by a listener in the situation of Figure 1.4.11 when the interchannel time difference is varied from 0 to +1 ms (left channel earlier) and –1 ms (right channel earlier). The curves show the results using different sounds: (*a*) dashed line = speech, solid line = impulses; (*b*) tone bursts; (*c*) continuous tones. (*From* [17]. *Used with permission.*)

intensity trading, and it is the justification for the balance control on home stereo systems, supposedly allowing the listener to sit off the axis of symmetry and to compensate for it by introducing an interchannel amplitude bias. There are some problems, however, the first one being that the trading ratio is different for different sounds, so that the centering compensations do not work

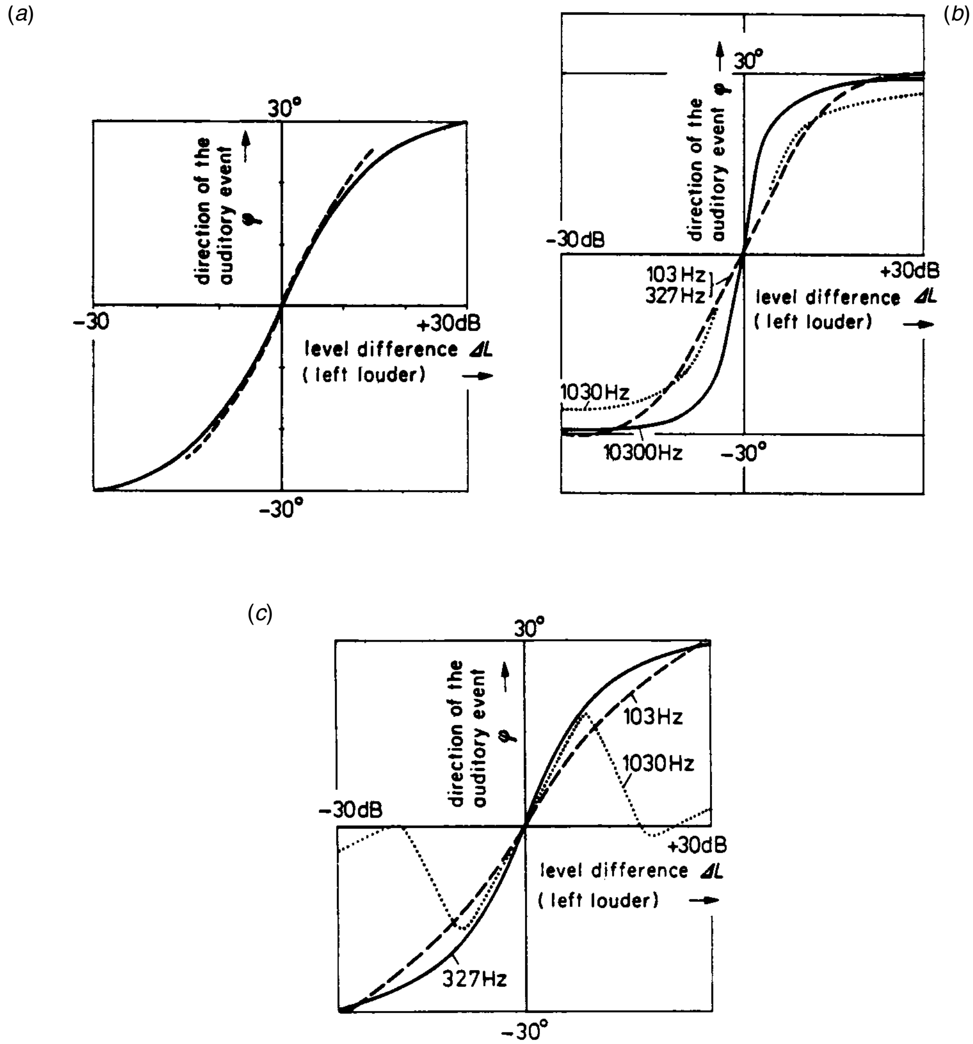


Figure 1.4.14 Direction of auditory images perceived by a listener in the situation of Figure 1.4.11 when the interchannel amplitude difference is varied from 0 to +30 dB (left louder) and –30 dB (right louder). The curves show the results with different sounds: (*a*) dashed line = speech, solid line = impulses; (*b*) tone bursts; (*c*) continuous tones. (*From* [17]. *Used with permission.*)

equally for all components of a complex signal; the image becomes blurred. The second problem arises when the listener moves beyond the limited range discussed previously, the simple form of summing localization breaks down, and the more complicated precedence effect comes into effect. In this region, it is to be expected that the auditory image will become rather muddled,

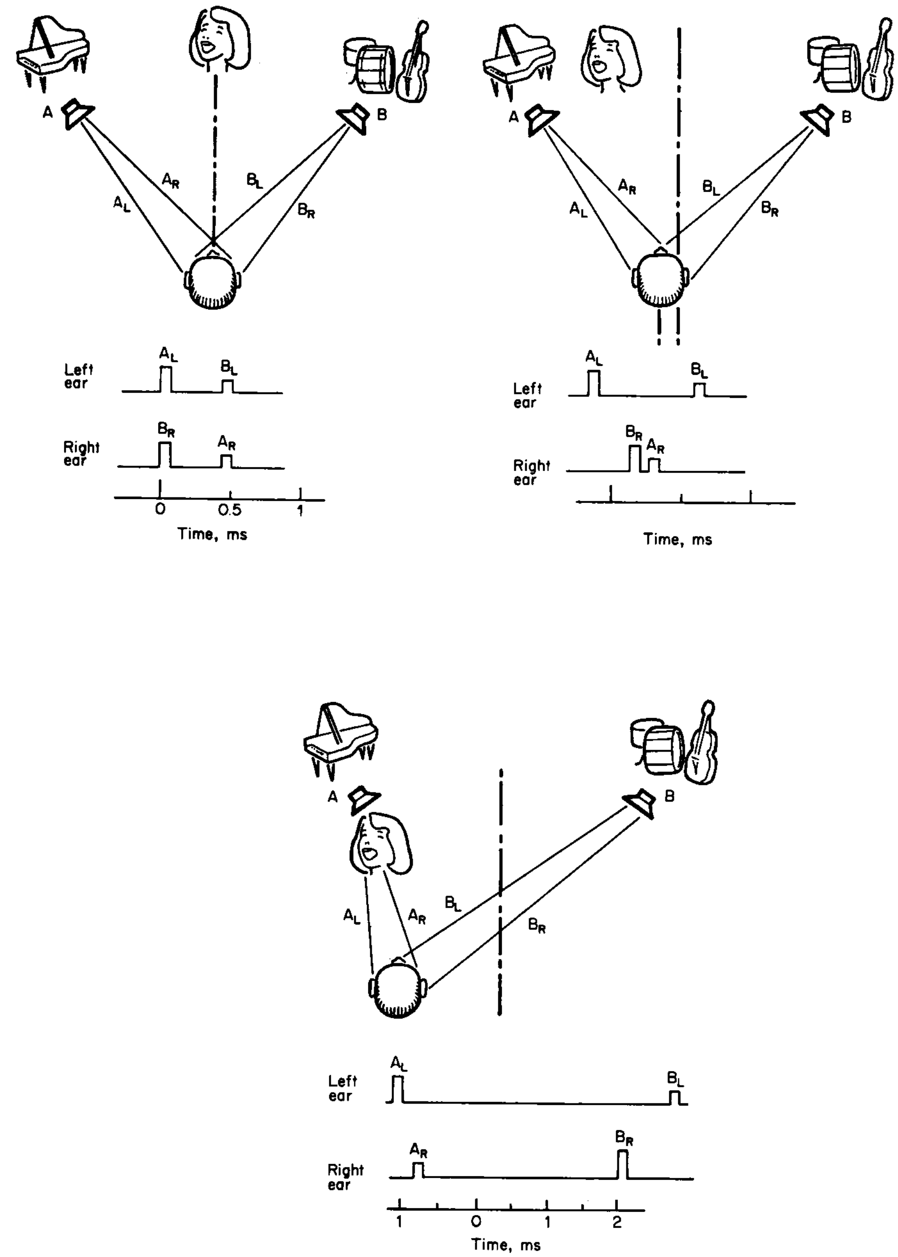


Figure 1.4.15 Sequence of events as a listener moves progressively away from the axis of symmetry in stereophonic listening.

increasing in size and spaciousness. Localization will tend to be related to the center of gravity of a spatially diffuse auditory event rather than of a specific compact event. Nevertheless, in recordings of ensembles with natural ambience the trading may be judged to be satisfactory, since the initial effect is, by design, rather diffuse. As the listener moves about, there will also be progressive changes to the timbre due to the directional properties of the loudspeakers and wave interference between essentially similar sounds arriving at the ears with various time delays.

Stereo Image Quality and Spaciousness

The position of auditory events is but a part of the total spatial impression. In stereo reproduction as in live performances, listeners appreciate the aspect of spaciousness as long as it creates a realistic impression. The process by which an impression of spaciousness is generated in stereo is much the same as in normal hearing—a reduction in interaural cross correlation. The tradeoff is also similar, in that as the feeling of space increases, the width of the auditory images also increases [33]. The extent to which the interchannel cross-correlation coefficient is altered to manipulate these effects is, therefore, a matter of artistic judgment depending on the type of music involved.

Special Role of the Loudspeakers

In the production of stereophonic recordings the impressions of image position, size, and spaciousness are controlled by manipulating the two-channel signals. However, the impressions received by the listener are also affected by the loudspeakers used for reproduction and their interaction with the listening room.

The directionality of the loudspeakers and the location of reflecting room boundaries together determine the relative strengths of the direct, early-reflected, and reverberant sounds that impinge on the listener. To the extent that the reflected sounds can reduce the correlation between the sounds at the two ears, it is clear that loudspeakers with substantial off-axis sound radiation can enhance the sense of spaciousness. For this to be effective, however, the listening room boundaries must be sound-reflecting at least at the points of the first reflections, especially the wall (lateral) reflections.

There is evidence that listeners in domestic situations prefer a certain amount of locally generated spaciousness [19, 34, 35]. In part this may be due to the more natural spatial distribution of the reflected sounds in the listening room as opposed to the recorded ambient sounds which are reproduced only as direct sounds from the loudspeakers. Loudspeakers placed in a room where the early reflections have been absorbed or directional loudspeakers placed in any type of room would be expected to yield a reproduction lacking spaciousness. This, it seems, is preferred by some listeners at home and many audio professionals in the control room [35, 36] especially with popular music. The fact that opinions are influenced by the type of music, individual preferences, and whether the listening is done for production or for pleasure makes this a matter for careful consideration. Once selected, the loudspeaker and the room tend to remain as fixed elements in a listening situation.

1.4.4 Sound in Rooms: The General Case

Taking the broadest view of complex sound sources, we can consider the combination of real sources and their reflected images as multiple sources. In this way, it is possible to deal with situations other than the special case of stereophonic reproduction.

1.4.4a Precedence Effect and the Law of the First Wavefront

For well over 100 years it has been known that the first sound arrival dominates sound localizaton. The phenomenon is known as the *law of the first wavefront* or the *precedence effect*. With time delays between the first and second arrivals of less than about 1 ms we are in the realm of simple summing localization. At longer delays the location of the auditory event is dictated by the location of the source of the first sound, but the presence of the later arrival is indicated by a distinctive timbre and a change in the spatial extent of the auditory event; it may be smeared toward the source of the second sound. At still longer time delays the second event is perceived as a discrete echo.

These interactions are physically complex, with many parametric variations possible. The perceived effects are correspondingly complex, and—as a consequence—the literature on the subject is extensive and not entirely unambiguous.

One of the best-known studies of the interaction of two sound events is that by Haas [37], who was concerned with the perception and intelligibility of speech in rooms, especially where there is sound reinforcement. He formed a number of conclusions, the most prominent of which is that for delays in the range of 1 to 30 ms, the delayed sound can be up to 10 dB higher in level than the direct sound before it is perceived as an echo. Within this range, there is an increase in loudness of the speech accompanied by "a pleasant modification of the quality of the sound (and) an apparent enlargement of the sound source." Over a wide range of delays the second sound was judged not to disturb the perception of speech, but this was found to depend on the syllabic rate. This has come to be known as the *Haas effect*, although the term has been extensively misused because of improper interpretation.

Examining the phenomenon more closely reveals a number of effects related to sound quality and to the localization dominance of the first-arrived sound. In general, the precedence effect is dependent on the presence of transient information in the sounds, but even this cannot prevent some interference from reflections in rooms. Several researchers have noted that high frequencies in delayed sounds were more disturbing than low frequencies, not only because of their relative audibility but because they were inclined to displace the localization. In fact, the situation in rooms is so complicated that it is to be expected that interaural difference cues will frequently be contradictory, depending on the frequency and temporal envelope of the sound. There are suggestions that the hearing process deals with the problem by means of a running plausibility analysis that pieces together evidence from the eyes and ears [38]. That this is true for normal listening where the sound sources are visible underlines the need in stereo reproduction to provide unambiguous directional cues for those auditory events that are intended to occupy specific locations.

1.4.4b Binaural Discrimination

The *cocktail-party effect*, in which it is demonstrably easier to carry on a conversation in a crowded noisy room when listening with two ears than with one, is an example of *binaural discrimination*. The spatial concentration that is possible with two ears has several other ramifications in audio. Reverberation is much less obtrusive in two-eared listening, as are certain effects of isolated reflections that arrive from directions away from that of the direct sound. For example, the timbral modifications that normally accompany the addition of a signal to a time-delayed duplicate (comb filtering) are substantially reduced when the delayed component arrives at the listener from a different direction [39]. This helps to explain the finding that listeners frequently enjoy the spaciousness from lateral reflections without complaining about the coloration. In this connection it has been observed that the disturbing effects of delayed sounds are reduced in the presence of room reverberation [37] and that reverberation tends to reduce the ability of listeners to discriminate differences in the timbre of sustained sounds like organ stops and vowels [20].

1.4.5 References

1. Shaw, E. A. G.: "The Acoustics of the External Ear," in W. D. Keidel and W. D. Neff (eds.), *Handbook* of *Sensory Physiology*, vol. V/I, *Auditory System,* Springer-Verlag, Berlin, 1974.
2. Fletcher, H., and W. A. Munson: "Loudness, Its Definition, Measurement and Calculation," *J. Acoust. Soc. Am*., vol. 5, pp. 82–108, 1933.
3. Robinson, D. W., and R. S. Dadson: "A Redetermination of the Equal-Loudness Relations for Pure Tones," *Br. J. Appl. Physics*, vol. 7, pp. 166–181, 1956.
4. International Organization for Standardization: *Normal Equal-Loudness Contours for Pure Tones and Normal Threshold for Hearing under Free Field Listening Conditions*, Recommendation R226, December 1961.
5. Tonic, F. E.: "Loudness—Applications and Implications to Audio," *dB,* Part 1, vol. 7, no. 5, pp. 27–30; Part 2, vol. 7, no. 6, pp. 25–28, 1973.
6. Scharf, B.: "Loudness," in E. C. Carterette and M. P. Friedman (eds.), *Handbook* of *Perception*, vol. 4, *Hearing*, chapter 6, Academic, New York, N.Y., 1978.
7. Jones, B. L., and E. L. Torick: "A New Loudness Indicator for Use in Broadcasting," *J. SMPTE*, Society of Motion Picture and Television Engineers, White Plains, N.Y., vol. 90, pp. 772–777, 1981.
8. International Electrotechnical Commission: *Sound System Equipment*, part 10, *Programme Level Meters*, Publication 268-1 0A, 1978.
9. Zwislocki, J. J.: "Masking—Experimental and Theoretical Aspects of Simultaneous, Forward, Backward and Central Masking," in E. C. Carterette and M. P. Friedman (eds.), *Handbook of Perception,* vol. 4, *Hearing*, chapter 8, Academic, New York, N.Y., 1978.
10. Ward, W. D.: "Subjective Musical Pitch," *J. Acoust. Soc. Am*., vol. 26, pp. 369–380, 1954.
11. Backus, John: *The Acoustical Foundations of Music,* Norton, New York, N.Y., 1969.

12. Pierce, John R.: *The Science of Musical Sound*, Scientific American Library, New York, N.Y., 1983.
13. Gabrielsson, A., and H. Siogren: "Perceived Sound Quality of Sound-Reproducing Systems," *J. Aoust. Soc. Am.*, vol. 65, pp. 1019–1033, 1979.
14. Toole, F. E.: "Subjective Measurements of Loudspeaker Sound Quality and Listener Performance," *J. Audio Eng. Soc.*, vol. 33, pp. 2–32, 1985.
15. Gabrielsson, A., and B. Lindstrom: "Perceived Sound Quality of High-Fidelity Loudspeakers." *J. Audio Eng. Soc.*, vol. 33, pp. 33–53, 1985.
16. Shaw, E. A. G.: "External Ear Response and Sound Localization," in R. W. Gatehouse (ed.), *Localization of Sound: Theory and Applications*, Amphora Press, Groton, Conn., 1982.
17. Blauert, J: *Spatial Hearing*, translation by J. S. Allen, M.I.T., Cambridge. Mass., 1983.
18. Bloom, P. J.: "Creating Source Elevation Illusions by Spectral Manipulations," *J. Audio Eng. Soc.*, vol. 25, pp. 560–565, 1977.
19. Toole, F. E.: "Loudspeaker Measurements and Their Relationship to Listener Preferences," *J. Audio Eng. Soc.*, vol. 34, part 1, pp. 227–235, part 2, pp. 323–348, 1986.
20. Plomp, R.: *Aspects* of *Tone Sensation—A Psychophysical Study*," Academic, New York, N.Y., 1976.
21. Buchlein, R.: "The Audibility of Frequency Response Irregularities" (1962), reprinted in English translation in *J. Audio Eng. Soc.*, vol. 29, pp. 126–131, 1981.
22. Stevens, W. R.: "Loudspeakers—Cabinet Effects," *Hi-Fi News Record Rev.*, vol. 21, pp. 87–93, 1976.
23. Fryer, P.: "Loudspeaker Distortions—Can We Rear Them?," *Hi-Fi News Record Rev.*, vol. 22, pp. 51–56, 1977.
24. Batteau, D. W.: "The Role of the Pinna in Human Localization," *Proc. R. Soc. London*, B168, pp. 158–180, 1967.
25. Rasch, R. A., and R. Plomp: "The Listener and the Acoustic Environment," in D. Deutsch (ed.), *The Psychology of Music*, Academic, New York, N.Y., 1982.
26. Shaw, E. A. G., and R. Teranishi: "Sound Pressure Generated in an External-Ear Replica and Real Human Ears by a Nearby Sound Source," *J. Acoust. Soc. Am.*, vol. 44**,** pp. 240–249, 1968.
27. Shaw, E. A. G.: "Transformation of Sound Pressure Level from the Free Field to the Eardrum in the Horizontal Plane," *J. Acoust. Soc. Am.*, vol. 56, pp. 1848–1861, 1974.
28. Shaw, E. A. G., and M. M. Vaillancourt: "Transformation of Sound-Pressure Level from the Free Field to the Eardrum Presented in Numerical Form," *J. Acoust. Soc. Am.*, vol. 78, pp. 1120–1123, 1985.
29. Kuhn, G. F.: "Model for the Interaural Time Differences in the Azimuthal Plane," *J. Acoust. Soc. Am.*, vol. 62, pp. 157–167, 1977.

30. Shaw, E. A. G.: "Aural Reception," in A. Lara Saenz and R. W. B. Stevens (eds.), *Noise Pollution*, Wiley, New York, N.Y., 1986.

31. Toole, F. E., and B. McA. Sayers: "Lateralization Judgments and the Nature of Binaural Acoustic Images," *J. Acoust. Soc. Am.*, vol. 37, pp. 319–324, 1965.

32. Blauert, J., and W. Lindemann: "Auditory Spaciousness: Some Further Psychoacoustic Studies," *J. Acoust. Soc. Am.*, vol. 80, 533–542, 1986.

33. Kurozumi, K., and K. Ohgushi: "The Relationship between the Cross-Correlation Coefficient of Two-Channel Acoustic Signals and Sound Image Quality," *J. Acoust. Soc. Am.*, vol. 74, pp. 1726–1733, 1983.

34. Bose, A. G.: "On the Design, Measurement and Evaluation of Loudspeakers," presented at the 35th convention of the Audio Engineering Society, preprint 622, 1962.

35. Kuhl, W., and R. Plantz: "The Significance of the Diffuse Sound Radiated from Loudspeakers for the Subjective Hearing Event," *Acustica*, vol. 40, pp. 182–190, 1978.

36. Voelker, E. J.: "Control Rooms for Music Monitoring," *J. Audio Eng. Soc.*, vol. 33, pp. 452–462, 1985.

37. Haas, H.: "The Influence of a Single Echo on the Audibility of Speech," *Acustica*, vol. I, pp. 49–58, 1951; English translation reprinted in *J. Audio Eng. Soc.*, vol. 20, pp. 146–159, 1972.

38. Rakerd, B., and W. M. Hartmann: "Localization of Sound in Rooms, II—The Effects of a Single Reflecting Surface," *J. Acoust. Soc. Am.*, vol. 78, pp. 524–533, 1985.

39. Zurek, P. M.: "Measurements of Binaural Echo Suppression," *J. Acoust. Soc. Am.*, vol. 66, pp. 1750–1757, 1979.

40. Hall, Donald: *Musical Acoustics—An Introduction*, Wadsworth, Belmont, Calif., 1980.

41. Durlach, N. I., and H. S. Colburn: "Binaural Phenemena," in *Handbook of Perception*, E. C. Carterette and M. P. Friedman (eds.), vol. 4, Academic, New York, N.Y., 1978.