5 The dimensions of sound

PREVIEW

The nature of the auditory experience provides the topic for this chapter. Now, with the knowledge of the physics of sound and the operation of the ear provided by the previous chapter, we can analyze our perception of sound and make some headway toward an explanation of that experience. The two major psychological dimensions of sound studied by psychologists have been loudness and pitch, and so this chapter starts there. In addition, we cover other phenomena, most especially the localization of sound sources. In all of this, we attempt to show through laboratory analyses how these sound experiences play a role in auditory perception in everyday life.

The musician may well be disappointed by the presentation of this chapter. To the musician, our discussion of pitch will seem too simple. Moreover, we barely discuss the perception of complex sounds, of sound quality, timbre, and the other attributes so important to the musical experience. The psychologist's notions of pitch and loudness differ somewhat from the musician's. Our terms refer to a very precisely defined sensation, in particular, to how a listener matches a complex sound to the loudness and pitch of a simple, pure sinusoidal sound. Thus, a psychologist speaks of a sound as having a single pitch and a single loudness. Musicians use these terms differently, often thinking of a sound as having several pitches and several loudnesses. A full treatment and understanding of the richness of the auditory experience and the quality of the sound does not yet exist. The procedures and issues discussed in this chapter provide the tools necessary to treat musical experiences in terms of our psychological understanding. More and more psychologists and musicians are starting to explore the world of sound. As they do so, our knowledge and our perceptual experiences should expand considerably.

In this chapter we talk about how acoustical information is interpreted. You need to know the concepts about the ear and about the specification of sound (frequency in *hertz* and intensity in *decibels*) presented in the previous chapter. If you do not, you should review that chapter.

Make sure you understand the diagram that shows the equiloudness contours and understand their implications for hearing (Figure 5-2). The loudness compensator on audio sets provides a good test. If you can explain the loudness compensator—why it is needed and how it works—then you probably have a good understanding of the sensitivity of the ear.

The concept of masking is important, for it is involved in a lot of real auditory phenomena, such as why the sounds of an airplane will drown out voices, but the equally loud sound of a flute playing some music will not.

The measurement of loudness is important. At least realize that loudness only tends to double when sound intensity gets ten times as large.

The sone scale is used to indicate how *loud* something is, and you should distinguish this from decibels, which tells how *intense* a sound is. Loudness is a psychological measure of sound. Intensity is a physical measure. Loudness depends upon sound intensity, frequency, and other variables. Do not confuse loudness and intensity.

In the discussion on pitch, again it is important to distinguish among several concepts: frequency is not the same as pitch. In addition, the psychological scale of pitch (mels) is not the same as the musical scale. Here is the start of the connection between what happens on the basilar membrane and in the neurons of the ear and what is perceived. A major purpose of this chapter is to consider how the mechanisms of the ear are related to our psychological perceptions.

Two of the most important points of this chapter are how the activity in the nerve fiber and the vibration of the basilar membrane affect perception, and how pitch is perceived. Note well the rationale for place theory and periodicity theory. First we teach you these theories, then tell you they are both incorrect. Don't despair. The theories are extremely important. Although there is a tendency to complain when someone tells you to learn something that is already known to be wrong, the pitch perception theories are important, useful, and not really that far wrong. We think it essential that you understand why each theory was proposed, why it was initially thought to be right, and what the problems are.

How sound is localized is the next topic, and again there are opposing theories. But both of the sound localization theories are considered to be correct, one working at low frequencies, the other at high frequencies. Understand the reasons why. These theories should help you understand stereophonic and quadrophonic sound reproduction.

SENSORY EXPERIENCES[1]

As we discussed in the beginning of Chapter 3, it is important not to confuse the psychological attributes of an experience with the physical attributes. The physics of sound can be specified with great precision. The psychological impressions that result from exposure to a particular physical sound are not so easily specified. Moreover, the psychological impressions may depend upon the history of experiences that the observer has had. With sound, the two most obvious psychological dimensions are loudness and pitch, but there are also other experiences of sound quality —timbre, dissonance, consonance, and musicality.

With the simplest of physical waveforms—a pure sine wave of the type produced by a simple, smooth whistle or an electronic oscillator—the physical aspects can be described by its frequency, intensity, and phase (when the signal starts). If the intensity of a sine wave is increased, the sound will increase in loudness. If the frequency is increased, the pitch will increase. As a result, it is tempting to associate the psychological dimensions of loudness and pitch with the physical dimensions of intensity and frequency; however, to do so is incorrect. For one thing, the relations are not linear: doubling the intensity does not double the loudness. For another, the relations are not independent: varying the frequency of a sound will affect both its loudness and its pitch. Finally, the relations are not constant: the perception of the pitch and loudness of a tone depends upon the context in which it appears, the nature of the other sounds that are present at the same time or slightly preceding the tone. Even the simplest physical dimension is subjected to a complex analysis by the nervous system.

It is important not to make the mistake of confusing the psychological perception of sound—loudness, pitch, and timbre—with the physical properties of intensity, frequency, and spectrum. They are different things. An orchestra creates a rich auditory experience. Rock groups, music synthesizers, electronic sounds are all combined in experimental works that stimulate the listener. New recording and playback techniques provide the means of recreating for the listener the experiences of the original event, be it a conference, a speech, a musical performance, or even the special effects imagined by creative composers that cannot exist, except through recordings.

Meanwhile, noise sources pollute the environment. The noises from aircraft annoy and disturb, sometimes being simply a tolerable nuisance, other times disrupting ongoing events, sometimes even causing physical damage and mental fatigue.

[1] This section is almost identical to the corresponding section in Chapter 3. If you read that section, you need only skim this one.

FIGURE 5-1

Table 5-1

Psychological Variables	Physical Variables: Of Primary Importance	Physical Variables: Of Secondary Importance
Hearing		
Loudness	Sound intensity	Frequency of sound waves (Hz)
Pitch	Frequency of sound waves (Hz)	Sound intensity
Timbre (quality)	Complexity of sound wave	—
Volume (size)	Frequency and intensity	—
Density	Frequency and intensity	—
Consonance (smoothness) Dissonance (roughness)	Harmonic structure	Musical sophistication
Noisiness	Intensity	Frequency composition, temporal parameters
Annoyance	Intensity	Frequency composition, meaningfulness

All these characteristics of sound are within the domain of the psychologist. From a knowledge of the mechanics of the ear, an understanding of the psychological dimensions of pitch and loudness, and a study of the phenomena of masking and auditory space perception, psychologists can talk about, explain, and predict many of the attributes of the auditory experience.

In this chapter, we examine some of the rich auditory experiences produced by sounds. We explore in turn four topics: loudness, pitch, the critical band, and auditory spatial perception. With each topic we introduce what has been learned from science and then deduce the practical implications. We examine the role of the four factors on the perceptions of music, speech, and noise.

LOUDNESS

The loudness of a tone depends upon both its intensity and its frequency. When frequency is constant, intense sounds appear louder than weak sounds. But when intensity is held constant, very high and very low frequency sounds seem much softer than sounds of intermediate frequency. In the extreme cases, this obviously must be true. Consider a whistle of intermediate frequency at a medium intensity level. Keep the intensity of the whistle constant but change the frequency so that it goes below 20 Hz or above 20,000 Hz (you have to do this electronically—you cannot do it by whistling). At these extreme frequencies, the sound becomes inaudible. Loudness, then, depends upon frequency, if for no other reason than the simple fact that there are limits in the range of frequencies to which the ear can respond. But loudness also depends on frequency within the normal hearing range.

Equiloudness contours

The interaction between frequency and intensity in the perception of loudness can be determined by asking people to compare two tones that have differing frequencies and intensities. Let one tone be the *standard* and give it a fixed frequency, intensity, and duration: for example, let the standard be a 1000-Hz tone with a sound pressure level (spl) of 40 dB spl presented for 0.5 sec. Let the second tone be the *comparison* tone. Make it 0.5-sec long also, but with a frequency different from that of the standard, say 3000 Hz. Now, the task of the listener is to listen alternately to the standard and the comparison tones, adjusting the sound level of the comparison until it sounds exactly as loud as the standard. When that is done, set the comparison tone to a different frequency and repeat the whole procedure. The typical result is shown in Figure 5-2: a curve describing the sound levels at which tones of various frequencies have the same perceived loudness as the standard. This curve is called an *equiloudness contour*. The level of the standard tone can be called the *loudness level* of the entire curve because it has been constructed by varying the comparison-tone frequency systematically while keeping the loudness of the comparison equal to the loudness of the standard.

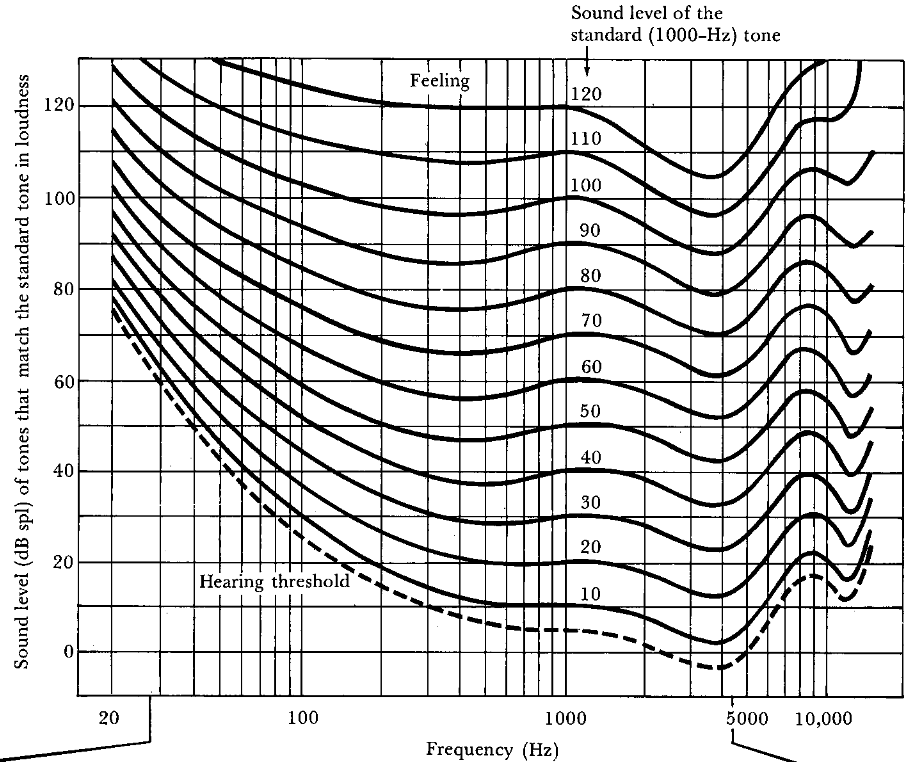
Figure 5-2 shows the results of an experiment using many different standards. For each curve, the standard tone was always at a frequency of 1000 Hz, but with different sound levels. The curve labeled 40 corresponds to the example just discussed: The standard tone had an intensity of 40 dB spl and a frequency of 1000 Hz. The curve labeled 100 is an equiloudness contour obtained when different frequencies are compared against a standard tone with an intensity of 100 dB spl, but still with a frequency of 1000 Hz.

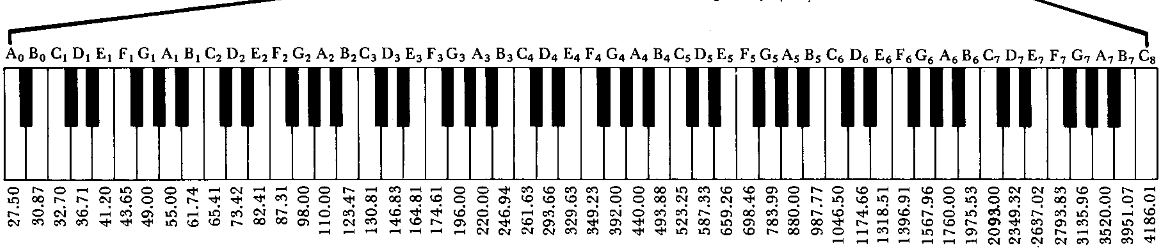
Note that Figure 5-2 shows how loudness varies with both sound level and frequency. Examine the curve in the figure labeled 50: the curve for a standard tone of 50 dB spl at 1000 Hz. When the comparison tone is around 20 Hz, it must be made quite intense (about 95 dB spl) in order to be just as loud as the 1000 Hz standard tone. As the comparison frequency is increased, the sound level of the comparison tone gets closer to the sound level of the standard. In fact, at 500 Hz, the comparison tone can be slightly weaker than the standard and still sound equally loud. Finally, although not shown on this figure, as the comparison tone goes above 10,000 Hz, it must be made more intense again to equal the standard tone in loudness.

The bottommost contour (the dashed line) shows the absolute sensitivity of the ear to different frequencies. Sounds below this line cannot

FIGURE 5-2

Equiloudness contours from data of Robinson and Dadson (1956).





be heard. Sounds on the line are just barely detectable (and are thus also assumed to be equal in loudness). At the other extreme, the topmost contour, very intense sounds lead, first, to a sensation of "tickling" in the ear and then, as intensity is increased further, to pain. Sounds in these regions can lead to ear damage. (Damage can also occur from sounds of weaker sound levels if there is prolonged exposure.) Note that the equiloudness contour at the threshold of pain is much flatter than the contour at the threshold of hearing. If sounds are intense enough, they tend to sound equally loud, regardless of frequency

Listening to music

It may come as a surprise to see where the sounds of the instruments of an orchestra lie on the equiloudness contours. The piano has the widest range of frequencies, going from about 30 to about 4000 Hz.

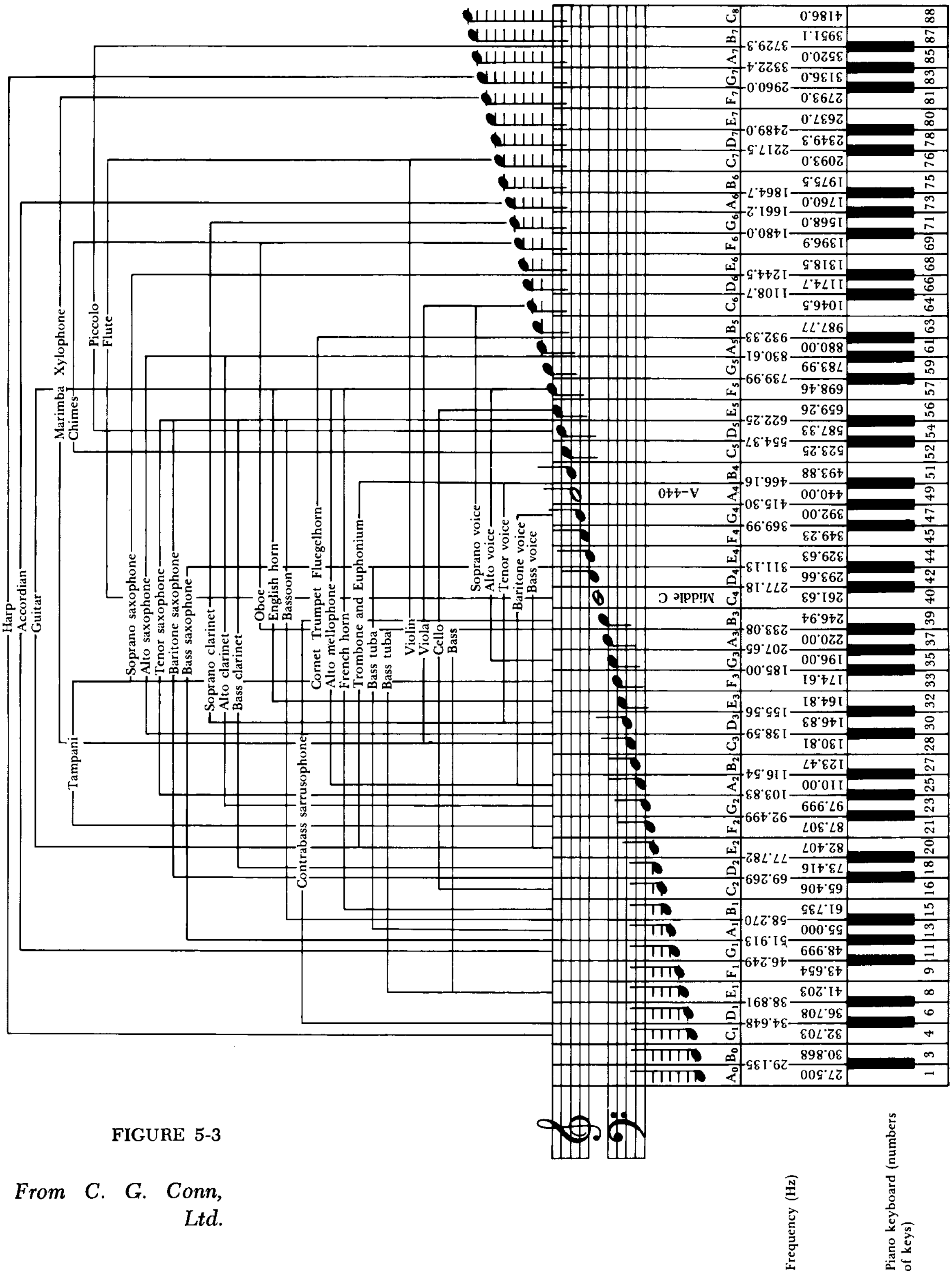


FIGURE 5-3

From C. G. Conn, Ltd.

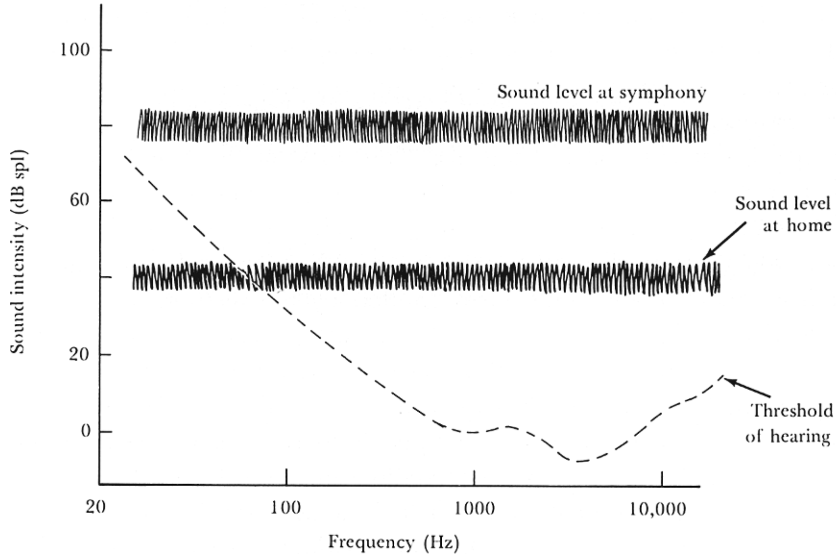
Middle C (C_4)[2] on the piano is about 260 Hz (261.63, to be exact).

To help you see where these frequencies lie on the equiloudness contours, we show a piano keyboard in Figure 5-2 and indicate what region it covers on the curves. Figure 5-3 shows the frequency range of a number of different musical instruments. Thus, most of the sounds of musical instruments lie in the region where the perception of loudness is most sensitive to changes in frequency. This has two results: One, unless you listen to the orchestra at reasonable intensity levels, you will not hear many of the frequencies emitted by the instruments; second, the loudness relations among the instruments, so carefully worked out and controlled by the conductor, depend upon your listening to the music at the same intensity that the conductor expected for the audience. When listening to recordings of symphonic music at home, you are not likely to turn up the level on your audio equipment to recreate the intensities originally present in the hall. You hear the music on a different set of loudness contours than the conductor planned for, so that you hear a different piece of music than the conductor intended.

The problems of playing back music so that it sounds the same in the home as it did when it was recorded are well known. Consider a segment of a symphony piece which is being played so that the overall sound intensity is approximately the same at different frequencies, as shown in Figure 5-4. When a recording is played back at home, the overall picture looks similar but the intensity is reduced. Now some of the sound levels fall below the threshold of hearing. Lower frequencies that were perfectly audible at the concert can no longer be heard. Moreover, as the levels change, the relative loudness at different frequencies also changes. If an organ plays a scale at very high intensity going from low notes to high ones, in the actual auditorium all the notes would be perceived to have approximately the same loudness (at high intensities, the equiloudness contours are relatively flat). At home, however, when the scale is played back at a reasonable listening level, not only would some of the lower frequencies be inaudible, but they would now fall in the region where frequency is affecting loudness: The notes would appear to get louder and louder up to about two or three octaves above middle C, where they would start decreasing in loudness. (Note from Figure 5-2 that as frequencies decrease below

[2] Subscripts indicate which octave of the note is referred to. The notation is the standard used by acousticians, but not always by musicians. The first C on the piano keyboard is named C_1. All notes within the octave immediately above it are given the subscript 1: D_1, E_1, . . . , B_1. The second C on the keyboard, and all the notes within the octave above it, are subscripted 2: C_2, . . . , B_2. By this scheme, the middle C on the piano keyboard is C_4; the note on which the instruments of an orchestra tune is A_4. The highest note on the piano is C_8; the lowest is A_0 (see Backus, 1968).

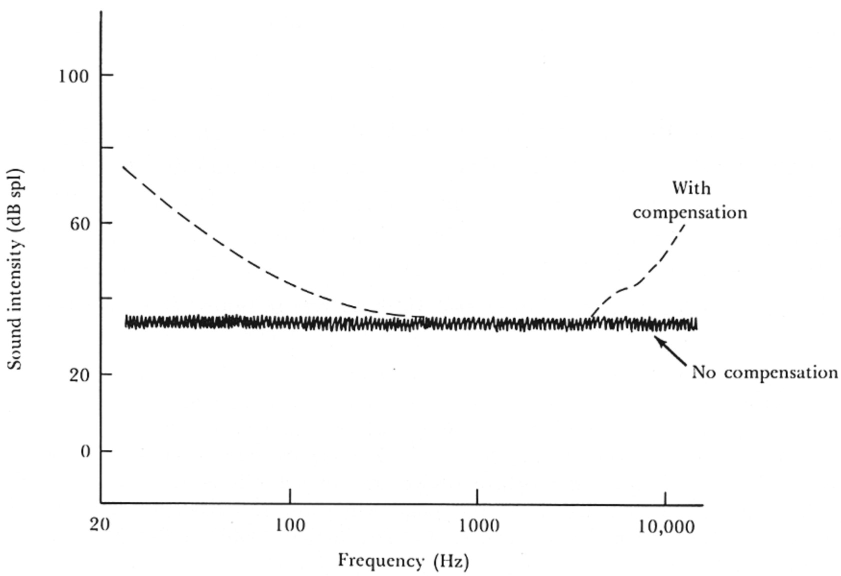
FIGURE 5-4



1000 Hz, they must be increased in intensity to be perceived as equally loud; thus, when intensity is constant, as frequency decreases the notes are perceived as getting softer.)

Loudness compensators. Most high-quality audio amplifiers now come with circuits that compensate for these psychological mechanisms. The control labeled *loudness compensator,* or sometimes simply *loudness,* makes the audio set overemphasize very low- and very high-frequency sounds when it is playing at low sound-levels. At high sound-levels, the loudness compensator should automatically be disconnected so that

FIGURE 5-5



it no longer has any effect. With good playback equipment, the effect of loudness compensation on the sound levels is shown in Figure 5-5 This results in a perceived loudness that looks like that shown in Figure 5-6. Actually, this compensation can work only if everything is done just right. The compensatory mechanism must take into account the peculiarities of the acoustics of the room where the speakers are located, as well as the particular sound equipment being used.

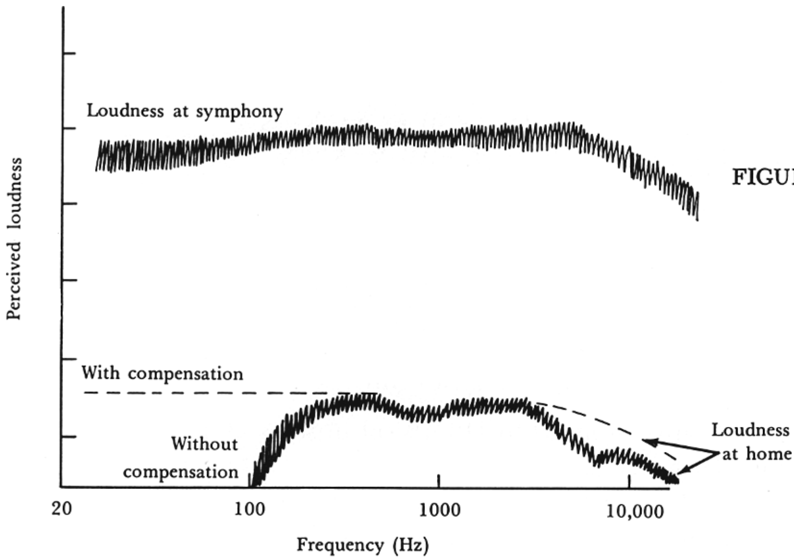


FIGURE 5-6

Masking

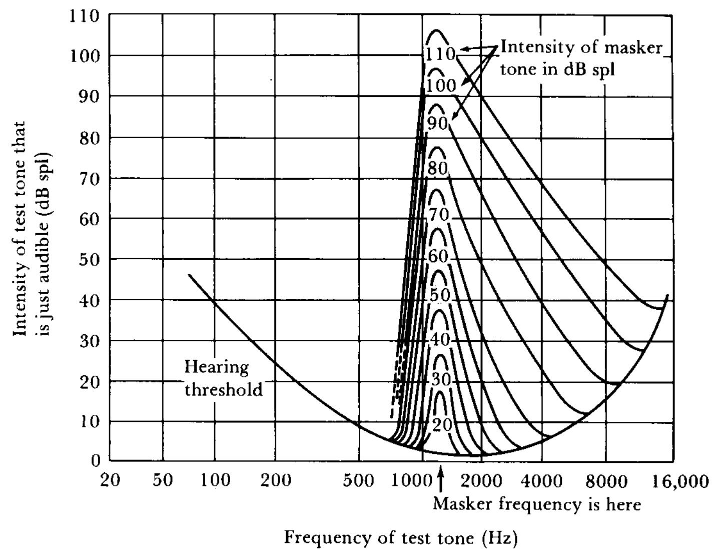
The loudness of a sound depends not only on its own intensity but also on other sounds present at the same time. Sounds *mask* one another: The presence of one sound makes another more difficult to hear. Rustling of papers, clapping of hands, coughing—all these tend to mask speech or music. To determine the effect of a masking, it is necessary to measure how much more intense a *test sound* must be in order to be heard in the presence of a *masking sound.* The procedure is basically similar to the one used for obtaining equiloudness contours.

The masking experiment. One way of doing the experiment goes something like this: Two tones are presented to an observer; one is called the *test tone,* the other the *masker.* The masker is set at some fixed intensity and frequency. Then the test tone is set to some fixed frequency and the intensity value adjusted until the test tone can just barely be detected. This procedure is repeated for different values of frequency of the test tone until an entire *masking curve* has been traced out, showing exactly how intense the test tone must be at different frequencies in order that it can be detected in the presence of the masker. Once the masking curve has been determined for one particular masker, the masker itself might be changed in either frequency or intensity and a new masking curve determined.

A typical result of this experiment is shown in Figure 5-7. In this case, the frequency of the masker[3] was held fixed at 1200 Hz and the

FIGURE 5-7

From Zwicker and Scharf (1965).



intensity was varied from 20 to 110 dB spl in steps of 10 dB to give 10 different masking curves.

The most striking feature of these data is the asymmetry. The masker has relatively little effect on tones below its own frequency of about 1200 Hz, but tones above this frequency are made much more difficult to hear by the presence of the masker.

The mechanism of masking. One of the explanations for this asymmetry comes from an examination of the vibration patterns of the basilar membrane. Remember the patterns shown in Figure 5-8: Low-frequency sounds tend to produce activity over much of the membrane, whereas high frequencies affect a more restricted region. If we examine these effects in more detail, as in Figure 5-9, we can compare the vibration patterns on the basilar membrane produced by the masker with those produced by the test tones. When the tone is weak and slightly higher in frequency than the masker, no part of the activity pattern produced by the tone manages to make itself felt above the pattern already caused by the masking noise. But the same weak tone at a frequency lower than the masker produces new activity in a separate nonoverlapping region. It does manage to be heard. Note that as the signal level is

[3] Actually, the masker was not a pure tone, but rather a narrow band of noise. Noise gives smoother results than does a tone. Other than this, the differences between the masking produced by a pure tone and by narrow-band noise are slight, and only of technical importance.

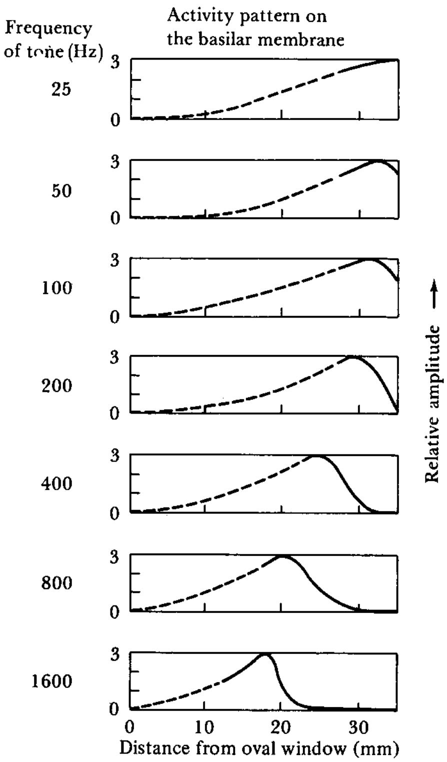


FIGURE 5-8

From Békésy (1949).

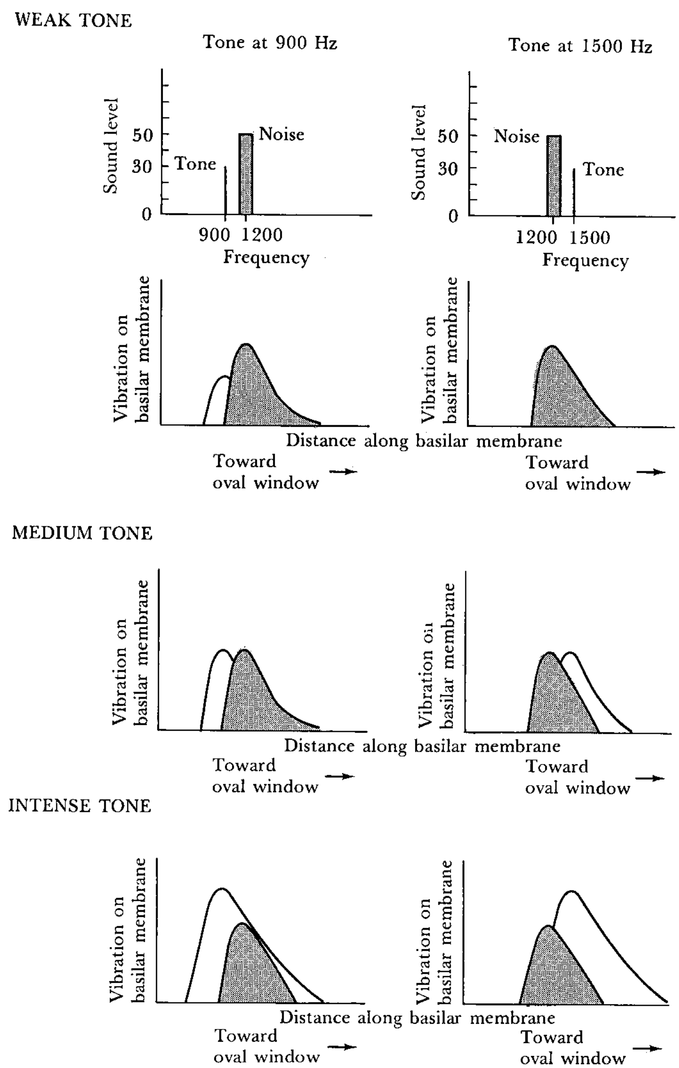


FIGURE 5-9

After Zwicker and Scharf (1965).

increased, the roles of the test signal and masker are reversed. A low frequency, relatively high-intensity test tone will mask the masker.

Masking of music. Masking adds another factor to the perception of loudness and of music. Intense, lower-frequency instruments mask the sounds of weak, higher-frequency instruments. The violas mask the violins; the timpani mask the violas; the brass mask the woodwinds. But when the sounds are played back in the home, the intensity is less than when they were recorded. As a result, the masking patterns are changed. Suddenly you can follow the fine fingerings of the violin or guitar, for the sounds of the basses are much reduced in level. Is this a virtue? Not necessarily. The composer, conductor, and the players did not have this in mind; their musical intuitions took into account the effect of masking and used it assuming it would be present for the listener. To eliminate the masking effect is to eliminate the sound balance among the instruments so carefully planned according to the grouping.

The measurement of loudness

Loudness measurements are of great importance for many practical problems. Since our psychological perception of loudness does not correspond directly to measures of physical intensity, it is essential to have methods that take these differences into account.

Sones. One such procedure is based on the method of *magnitude estimation.* We present a person with two tones, both, say, at 1000 Hz, and ask him how many times louder one sound appears to be over the other. The question is a peculiar one, but people can and do answer it sensibly. (For more information and some examples you can try yourself, see Appendix A.)

The results of the magnitude estimation procedure show that loudness increases as the cube root of sound intensity. That is, the psychological *J*udgment of loudness, *J*, is related to the physical *I*ntensity of the sound, *I*, by a power law of the form

$$J = kI^{0.3}\,.$$

This value of exponent (0.3) is very convenient. It works out that if the sound intensity is specified in decibels, a 10 dB increase always changes loudness by a factor of 2. Every time the physical intensity is multiplied by 10, psychological loudness is multiplied by 2, as shown in Figure 5-10.

This measurement procedure has been standardized by the International Standards Organization. The unit of loudness is the *sone.* By definition, the loudness of a 1000-Hz tone at an intensity of 40 dB spl is equal to 1 sone. Therefore, a 50 dB spl 1000-Hz tone would have a loudness of 2 sones, and a 100 dB spl 1000-Hz tone would have 64 sones.

To get the loudness of tones at other frequencies, the equiloudness contours can be used. All tones on the equiloudness contour in Figure 5-2 marked 40 have a loudness of 1 sone. Those on the contour marked 50 have a loudness of 2 sones, and on the contour marked 60 have a loudness of 4 sones. Each increase in contour by 10 dB doubles the sone value: Each decrease of 10 dB halves the number of sones.

The sone measurement describes the perceived loudness of pure tones. With complex sounds that contain many frequency components, such as voices, orchestras, or aircraft and automobile sounds, the loudness is determined by comparing them with a 1000-Hz standard. The sone value at which the 1000-Hz tone appears as loud as the complex sound is the sone level for that complex sound. The sone values for some typical sounds are shown in Figure 5-10.

Computation of loudness values in sones is fairly complex, but it is now possible to buy meters that give sone values directly. The meters incorporate computers inside them that combine the loudness values measured

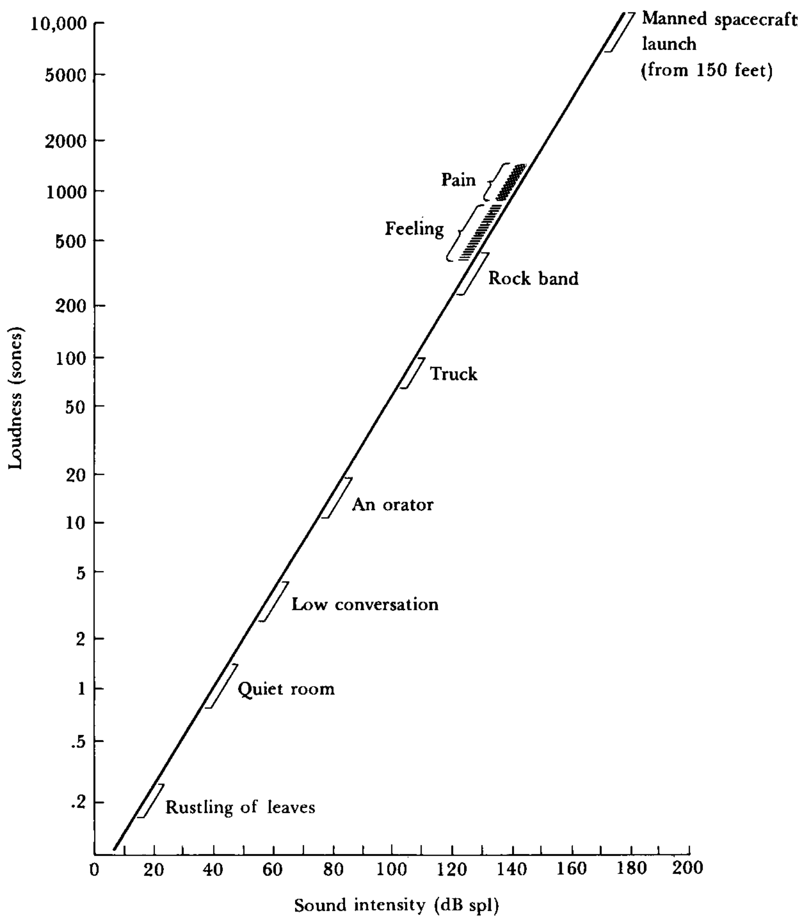


FIGURE 5-10

at different parts of the frequency range according to a formula based upon the differing sensitivity of the ear to different frequencies (as shown in Figure 5-2), as well as to the properties of the *critical band* (discussed later in this chapter). An illustration of the use of sones as a measure of loudness is shown in Figure 5-11, taken from *Consumer Reports* magazine. This figure shows a comparison of sone levels for the interior noise of automobiles when driven over different types of roads at different velocities: at 30 mph (around 50 km/hr), car 1 (the left column) is about 1½ times as loud as car 2 (the right column).

extre... gas mi...e between stop-and-go and open-road trips. Tank mileage is that actually obtained during a 195-mile trip in convoy. Constant-speed mileage is obtained by driving two ways on a substantially level road and averaging the results.

BRAKING. Minimum-distance controlled stops are shortest distance (to nearest five feet) achieved in several attempts, both without and with wheel-locking, within a 12-foot lane. (Note that wheel-locking may provoke directional instability.) Distances apply only to CU's test conditions, but the relative ranking should remain consistent under most conditions. Fade test includes 10 moderate stops at ⅓-mile intervals. Difference in pedal effort between first and 10th stops is degree of fade. Maximum Acceptable pedal effort is 150 pounds.

INTERIOR NOISE. As measured by CU. Sones are a directly proportional measure of loudness.

CONSUMER REPORTS

	Car 1	Car 2
¼ mi... speed at en... (mph)		
Passing: 45 to 65 mph (seconds)	9.5	11.5
FUEL ECONOMY		
TANK MILEAGE OBSERVED ON 195-MILE TEST TRIP (mpg)	29	19
RANGE OF GAS MILEAGE TO BE EXPECTED IN NORMAL USE (mpg)	19-32	12-22
CONSTANT-SPEED GAS MILEAGE (mpg)		
at 40 mph	42	24.5
at 50 mph	34	23
at 60 mph	29.5	20.5
LEVEL BRAKING FROM 60 MPH		
LEVEL BRAKING FROM 60 MPH		
Minimum-distance stop with no wheels locked (feet)	170	180
Minimum-distance stop with some or all wheels locked (feet)	170	150
FADE TEST: Pedal effort for initial ½-g stop (pounds)	55	55
Effort for 10th repeated stop (pounds)	85	70
INTERIOR NOISE		
smooth road at 30 mph (sones)	28	19
coarse road at 30 mph (sones)	38	26
highway at 60 mph (sones)	43	33

A *Prices as of May 5, 1975.*
B *Four-door model with several options.*
C *Price includes automatic transmission, power steering, and several other options.*

FIGURE 5-11 *Sone values provide a way of directly comparing the noise levels inside different automobiles under different driving conditions. (From Consumer Reports, with permission.)*

PITCH

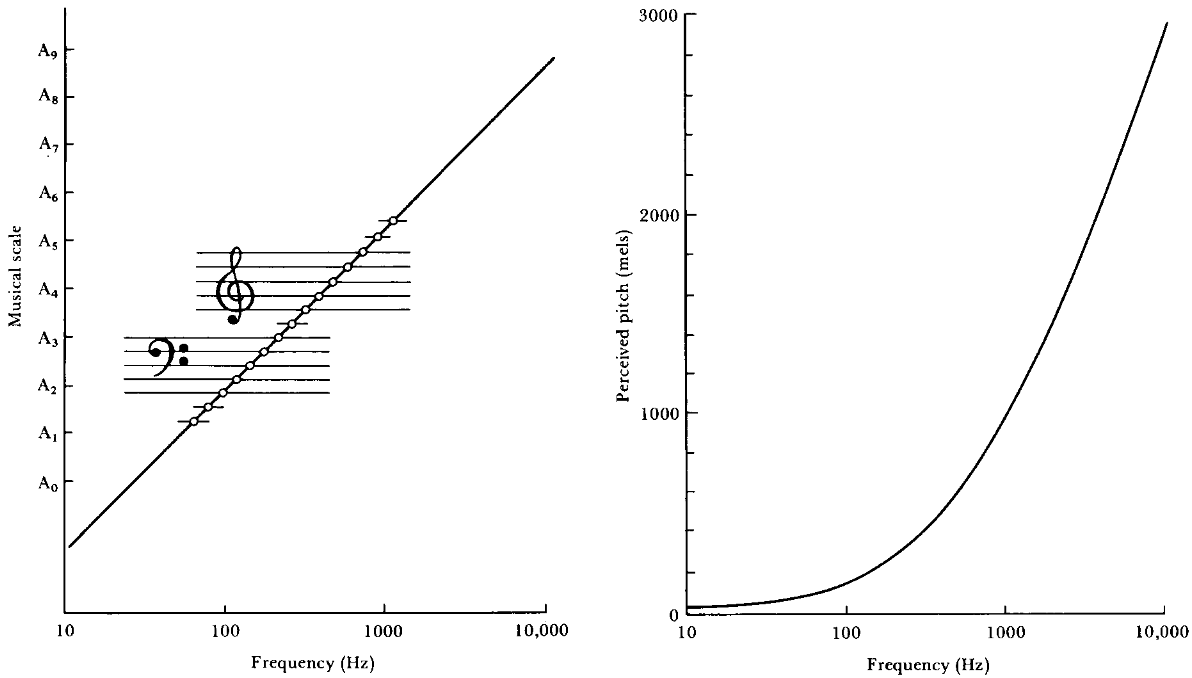
The musical scale

The musical scale of pitch is logarithmically related to sound frequency. Each octave in the standard musical scale is exactly twice the frequency of the previous octave. The note that orchestras use to tune their instruments, A_4 (the A above middle C), has a frequency of 440 Hz. The A's one and two octaves above that, A_5 and A_6, have frequencies of 880 and 1760 Hz, respectively. Similarly, the A's one and two octaves below have frequencies of 220 and 110 Hz, respectively. In an even-tempered musical scale, then, increasing the note by an octave doubles its frequency. Moreover, there are 12 equally spaced notes in an octave (counting all the whole and half notes). In order to divide the frequency range spanned by an octave into 12 equal intervals, each note is exactly $2^{1/12}$ times the frequency of the one before it.

The mel scale

Is the note of one octave perceived to be twice the pitch of the same note in the preceding octave? Our intuitions suggest yes, but experimental data indicate the answer is no. When subjects are presented with different notes and asked to judge the pitch relations among them, their perceived pitch does not follow the musical scale. Doubling or halving the frequency of the note does not double or halve its perceived pitch. (This result comes from use of the magnitude estimation procedure, described in Appendix A.) The actual relationship is shown in Figure 5-12. The unit of pitch in this diagram is called the *mel*. By definition, a tone of 1000 Hz (at 60 dB spl) has a pitch of 1000 mels.

Although this result may be incompatible with our intuition about pitch, it is highly compatible with some of the concepts in music composition. Musicians frequently debate the consequences of transposing a piece from one key to another. If a piece is written in C major and then transposed to A major, should it matter? If the change from one note to another always has the same psychological magnitude, regardless of the notes involved (equivalently, that raising a note an octave doubles the pitch), then why should it matter if the piece is transposed? The psychological distances between the notes will be the same, regardless of the key the piece is played in. But most musicians argue that transposition changes the character of the piece. The effect is subtle, but it is there. This argument is compatible with the psychological judgments



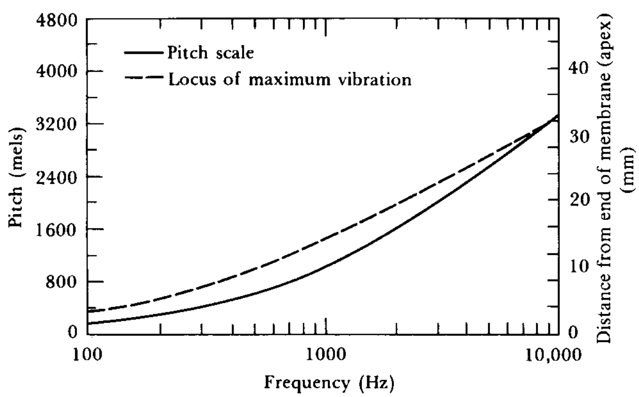
$mels = 2410 \log (1.6 \times 10^{-3} f + 1)$ FIGURE 5-12 of pitch relations. The change in perceived pitch involved in going from a C_4 to a D_4 is different from that involved in going from an F_4 to a G_4 or, for that matter, going from C_5 to D_5, an octave up.

Place theory: position on the basilar membrane

To consider the question of what determines our perception of pitch, let us again return to the vibration pattern along the basilar membrane. Different frequencies set up different patterns of activity on the membrane. The location of the maximum vibration moves systematically from the oval window end of the membrane toward the apex as the frequency goes from high to low. As early as 1863, the German physicist Helmholtz proposed that pitch is determined by the position of the maximum vibration along the membrane. Although Helmholtz' reasons were inaccurate, his conclusion was sound. Psychological distance between the pitches of two tones seems to be related to the physical distance between the position of the peak activity produced by the tones. The two functions are shown in Figure 5-13. Here, location of the maximum of the vibration patterns produced by tones of different frequencies is plotted in terms of its distance from the far end (the apex) of the membrane. Perceived pitch as measured by the mel scale is also shown. The two functions are similar, not identical.

FIGURE 5-13

From Zwislocki (1965).



Why should distance be the critical feature? It is of no intrinsic value to the nervous system unless there are neural mechanisms to take advantage of it. The way the 30,000 fibers of the acoustic nerve distribute themselves along the membrane suggests such a mechanism. Near the oval window and for the first few turns of the cochlea, the density of the neurons appears to be constant at approximately 1150 ganglion cell neurons per millimeter. But the density of neurons decreases toward the apex. If a higher density of neurons provides more precise position information, then this distribution suggests there should be less sensitivity to changes in activity patterns in the low-frequency region—the points nearer the apex. When the curve is corrected for the relative density of the neurons, there is very good agreement: Each unit change in pitch on the mel scale is approximately equal to a movement of the vibration pattern along the membrane by 12 neurons.

The exact sensitivity of the ear to changes in frequency can be measured directly by successively presenting pairs of tones to an observer who must decide whether they have the same or different frequencies. We obtain a measurement of the *j*ust *n*oticeable *d*ifference (the *jnd*) between frequencies. The ability to make such discriminations varies with frequency. At 100 Hz, a 3% change in frequency (3 Hz) is necessary before the perception of the sound is just noticeably different. Thus, at 100 Hz, the jnd for frequency is 3 Hz, or about 3% of the frequency. This percentage value steadily decreases until it reaches a minimum of around 0.2 to 0.3% at 1000 Hz. For low frequencies, the jnd for frequency is approximately constant in absolute value. Thereafter, the percentage change required to make the discrimination remains reasonably constant at about 0.3%.[4]

If the jnd for frequency is compared with the distance between the peaks in activity along the basilar membrane produced by the two frequencies being discriminated, there is good agreement in high-frequency regions but disagreement at low frequencies (Figure 5-14). But, as before, we should really take into account the way that the hair cells are spread out along the membrane. When the curve is corrected for the relative density of neurons, the match is improved: We can discriminate the difference between two frequencies whenever their peak activity is separated along the membrane by about 52 neurons.

An examination of the location of maximum vibration along the membrane, coupled with a consideration of the neural distribution, describes

[4] This pattern of changing discriminability is common to a number of different signal dimensions in different sensory systems. The discriminability between the intensities of an auditory signal, for example, shows a similar pattern with the jnd being approximately constant in absolute value for low intensity signals and then having an approximately constant **percentage** value as the intensities move into the middle regions of the hearing range. A constant percentage associated with discrimination is typical of many types of measuring instruments. It results from the fact that the variability in the measures being taken often depends on the level being measured, and increases with increasing levels. Consequently, the absolute size of the signal change needed for reliable discrimination will also increase. When these size increases result in a constant percentage change, the system is said to be following Weber's Law, named after the physiologist Weber (1795–1878), a contemporary of Helmholtz. If we let ΔI stand for the size of the jnd, the change in intensity that a signal must make in order for that change to be just noticeable, and I for the signal intensity, Weber's Law states that

$$\Delta I = kI,$$

where k is the relative change ($100k$ is the percent change).

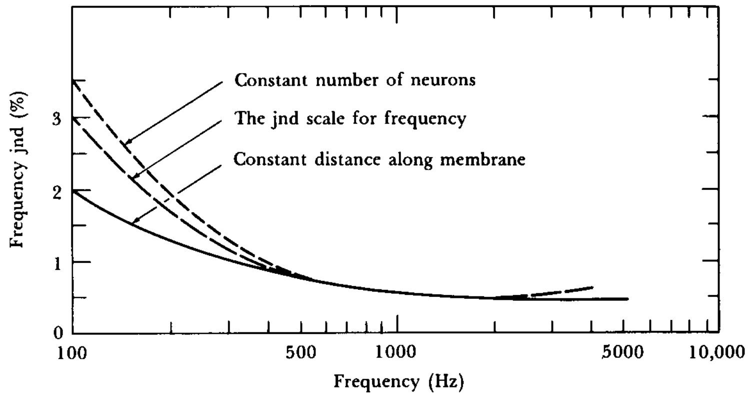


FIGURE 5-14

From Zwislocki (1965).

both the subjective perception of pitch and the ears' sensitivity to changes in frequency:

- 1 jnd is approximately 52 neurons;
- 1 mel is approximately 12 neurons.

Note that a pitch change of 1 mel is less than one jnd: It cannot be detected. Pitch must change by 4 or 5 mels before an observer can detect the change. At higher frequencies (above 500 or 1000 Hz), where the density of neurons is reasonably constant along the membrane, the jnd represents a constant distance—about 0.05 mm or about .002 inches—along the membrane.

Periodicity pitch

The analysis of the patterns of vibrations along the basilar membrane describes a number of the phenomena associated with our perception of pitch. But some puzzles remain.

Loudness and pitch. The ear is not very sensitive to low-frequency sounds, yet many musical instruments produce sound frequencies in this insensitive region. Thus, a note played softly will have much of its energy lying below the threshold of hearing. What happens to our perceptions? Obviously one note played more and more softly sounds much the same, but why? Shouldn't the same piano note keep changing in pitch as it gets softer and softer and as more and more of its low-frequency components become inaudible?

Consider the piano note C_3, the C below middle C—not a particularly low note. It has the same pitch as a tone of 131 Hz (actually 130.9 Hz). But the piano note is not a simple tone. Look at the spectrum shown in Figure 5-15. Although there is more energy at 131 Hz than at other frequencies, there actually is some energy over a wide range of frequencies. As the note is played more and more softly, these low-frequency components will drop in intensity below the level that can be

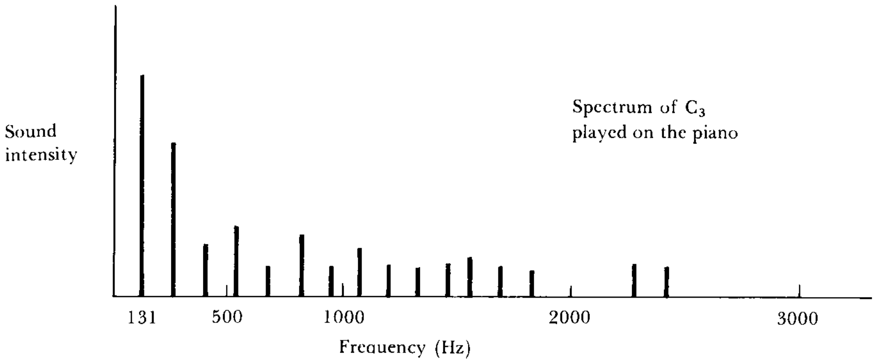


FIGURE 5-15

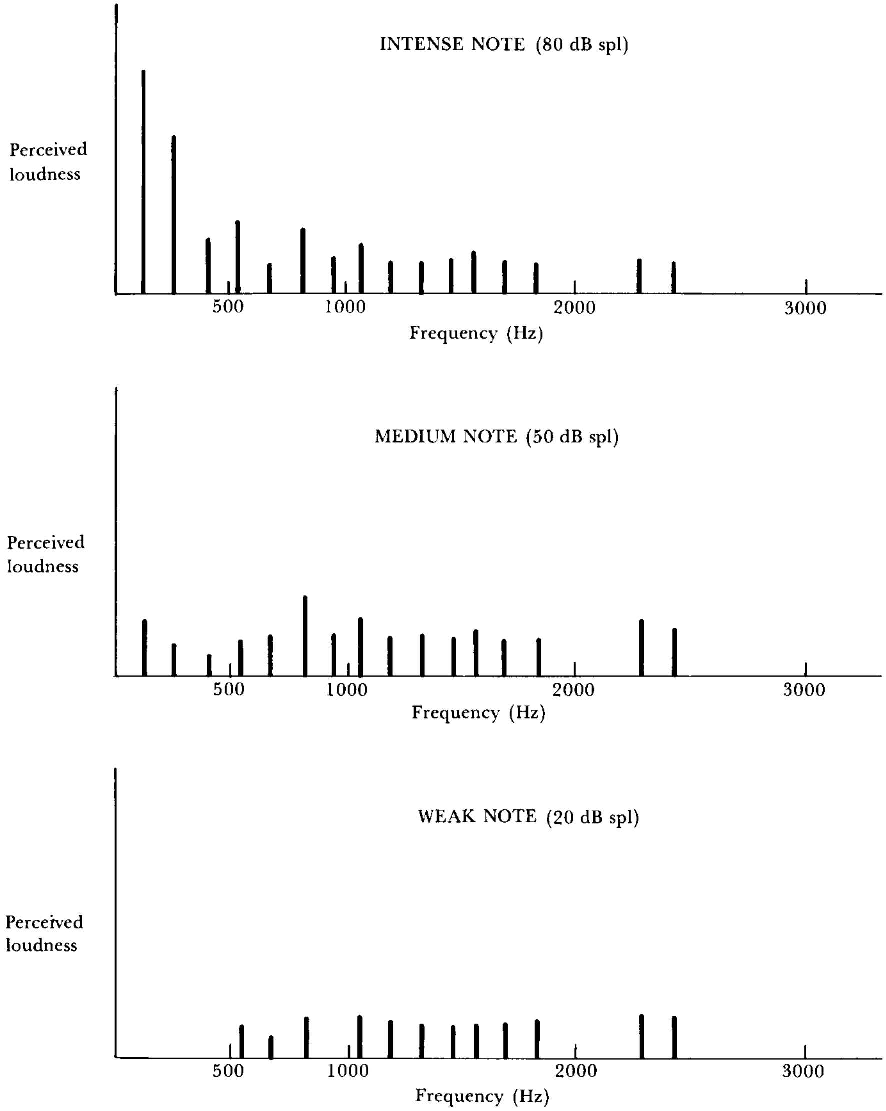
heard. Thus, the lowest frequency present will change from 131 Hz, when the note is played at a comfortable level, to 262, to 393, and finally to 524 Hz, when the note is played reasonably softly. A pure tone with a frequency of 524 Hz matches C_5 in pitch. This is a reasonably high pitched note—it is one octave above middle C. But quite clearly, there is something peculiar going on here. A musical note simply does not appear to change in pitch in this manner as it gets softer. If pitch is determined by the location along the basilar membrane, why does the pitch of a complex tone, such as a piano note, appear to remain constant even though its frequency structure is changing? How can the piano note continue to have the pitch of its fundamental frequency of 131 Hz when the lowest frequency that is audible is 524 Hz? How do we hear the missing fundamental?

Figure 5-16

The case of the missing fundamental. To answer the question, consider a simpler situation, shown in Figure 5-17. Two pure tones, 1000 Hz (upper row) and 1100 Hz (middle row) are added to produce a complex waveform (lower row). Note that even though the only sound energy present in the system is at the two frequencies 1000 and 1100 Hz, the resulting wave pattern appears to vary at an overall rate of 100 Hz. The phenomenon in the diagram is called a *beat*. Two sinusoidal waves played together produce a beat pattern, a regular rise and fall in the sound energy at a rate equal to the difference of the frequencies of the component sine waves. On the basilar membrane, however, this beat component should not be present. There are no physical frequencies in the wave corresponding to the beat frequency. Maximums in the activity pattern should be produced only at the locations corresponding to the actual physical components of the sound—1000 and 1100 Hz. Thus, if we perceive pitch by noting the points of maximum vibration, we should perceive the 1000-Hz and 1100-Hz components, but not the 100-Hz beat frequency. In fact, we do hear the beat.

There are two possible theories to explain this phenomenon. One proposes that the perception of the beat frequency results from the

FIGURE 5-16



fact that the ear is an imperfect transmitter of sound: It has *nonlinearities*. The mechanical structure in the ear (mainly the middle ear) actually adds extra frequencies to the incoming acoustic signal. In particular, the beat frequency is added, and thus the basilar membrane is activated in this frequency region. This explanation is compatible with the notion that a particular pitch is only perceived when there is a corresponding maximum in the vibration pattern on the membrane: the *place theory* of pitch perception.

The second theory emphasizes the importance of the synchronized firing of neurons to the changing pressures in the acoustic wave. The nerve cells reacting to the activity patterns on the basilar membrane

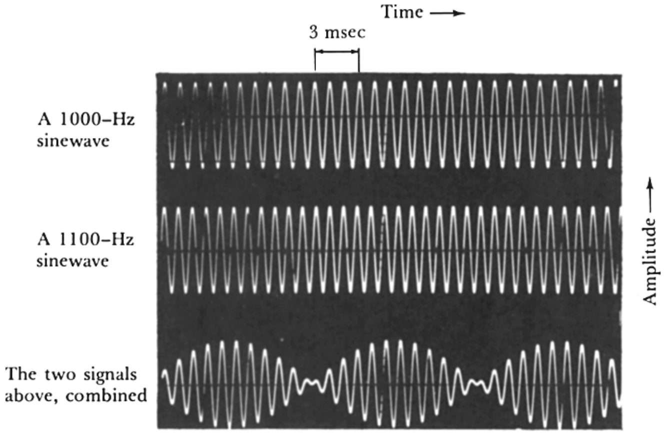


FIGURE 5-17

Beats. The fluctuation in overall sound pressure which results when two sine waves of different frequencies are added together. The resulting pattern fluctuates ("beats") with a frequency given by the difference between the frequencies of the components. In this illustration, the 1000-Hz sine wave is mixed with a 1100-Hz sine wave to give a beat frequency of 100 Hz. Thus, the beat pattern repeats itself every 10 msec. Note that there is no sound energy present at a frequency of 100 Hz. We did not create a new sine wave by adding together the 1000- and 1100-Hz signals: Simply the overall envelope of sound pressures varies at the beat frequency.

fire in synchrony with the regular rise and fall of the beat frequency. This synchronization in the neural responses is at the basis of the perceived pitch: the *periodicity theory* of pitch perception. These are the two major explanations of pitch perception and the study of the missing fundamental is the key to evaluating them. In the next section we present an experiment that helps distinguish the two theories.

Masking the missing fundamental

Consider a complex sound made up of the following frequencies:

1000 Hz   1200 Hz   1400 Hz   1600 Hz
1800 Hz   2000 Hz   2200 Hz

If subjects are asked to adjust an oscillator so that its pitch is the same as that of the complex sound, they will set the oscillator to 200 Hz.[5] Both theories of pitch perception can explain this simple phenomenon.

The place theory explanation. The ear is nonlinear. It produces difference frequencies: The seven tones presented give six opportunities for a 200-Hz difference frequency to be added by the distortion introduced

[5] The experiment described in this section was performed by Patterson (1969).

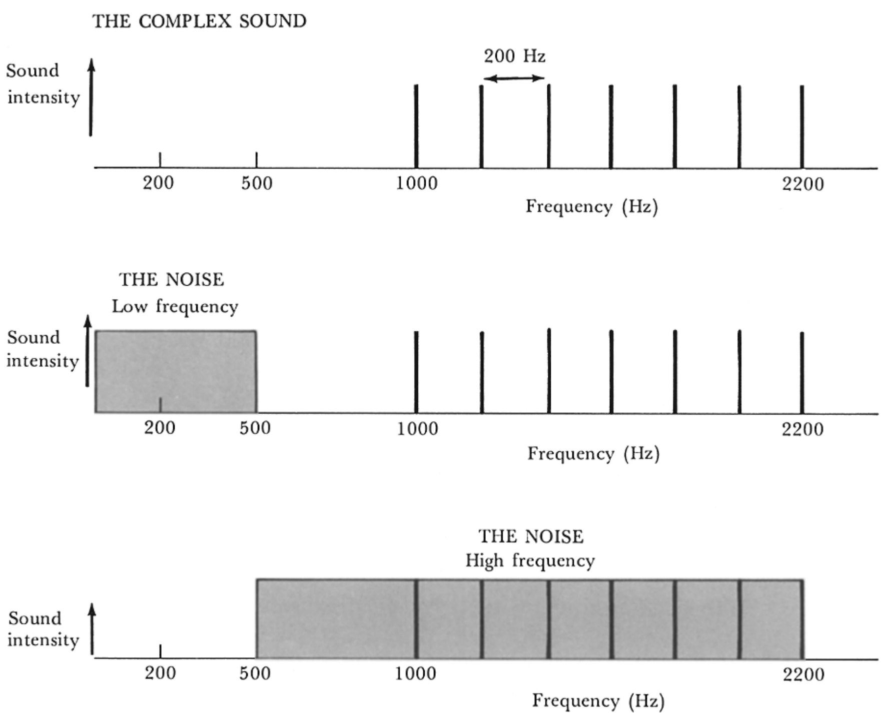
during transmission. This difference frequency, then, is a prominent contributor to perceiving the pitch of the sound.

The periodicity pitch explanation. The overall impulse flow in the auditory nerve is following the beat pattern of the sounds. There are nonlinearities. But neural activity is rising and falling regularly 200 times per second, and it is this activity pattern which is at the basis of the perceived pitch of the sound.

The critical difference, then, is whether the membrane is actually being activated in the 200-Hz region. The place theorist thinks that the subject perceives the basilar membrane vibrating at the 200-Hz location. The periodicity pitch theorist believes that the membrane is vibrating only in the high-frequency region between 1000 and 2200 Hz and that the nerve firings at these locations are synchronized to 200 Hz. What is the critical experiment? Disrupt the membrane in the location around 200 Hz and find out if the subject can still hear the associated pitch.

It is rather easy to disrupt the membrane in this way. One method is to add low-frequency noise to the signal—a sound that contains energy at all frequencies below some value. To be sure the noise will mask out any low-frequency activity on the membrane, add noise that contains all frequency components up to 500 Hz. How intense should the noise be?—intense enough to disrupt a real tone, if one were there. To determine this value, first present a real 200-Hz tone and have the subject adjust its

FIGURE 5-18



intensity so that it sounds exactly as loud as the 200-Hz tone in the complex sound. Then add low-frequency noise in an amount sufficient to mask completely the real 200-Hz tone. This noise level should be sufficient to mask any activity produced by nonlinearities. Once again, the complex tone is turned on. This time the masking noise is added. Will the 200-Hz component still be audible?

The periodicity theorist is correct: The missing fundamental is still heard when all activity in its frequency region is being physically masked by noise. Moreover, to prove that the noise does have the proper masking effect when placed appropriately, a high-frequency noise can be tested —a noise containing all frequencies above 500 Hz. The neural response pattern will be disrupted by the high-frequency noise, and the missing fundamental should no longer be heard. The result? With high-frequency noise presented, the low-frequency fundamental is no longer audible.

PITCH DISCRIMINATION WITHOUT A BASILAR MEMBRANE[6]

One way to test for the workability of a periodicity theory is to study an animal that has no basilar membrane: Can it discriminate different audio frequencies? According to place theory, if there is no basilar membrane, there can be no encoding of pitch. According to periodicity theory, pitch discrimination should be reasonably good, at least up to the point where the neural firing can no longer keep up with the signal.

The goldfish is such an animal. It has hair cells, but no membrane. The ears of most fish are rather different from those of mammals—for good reason. Not only are they evolutionarily less advanced, but the water makes a peculiar medium for sound. Water is more dense than air: Sound travels five times as fast in water as in air. Moreover, the density of the water does not differ much from the density of the body tissues and fluids. This means that the entire outer ear and middle ear are unnecessary, perhaps even harmful. Sound tends to travel right through fish, with no diminution in intensity. The ears of fish are located at the air bladders, and the distinctions found among the ears of fish seem related to the distinctions in the way their air bladders are located. What is most important, of course, is that fish—and the goldfish in particular—do have hair cells and acoustic nerves, but no basilar membrane. How, then, can they discriminate frequency?

When the goldfish is properly trained, it can indeed tell one frequency from another. The test is to hold the fish securely in a cheesecloth-padded harness and continually pair the presentation of a tone with an electric shock. Figure 5-19
The fish is presented with a series of tones of the same frequency. Then one tone is changed in frequency, and a shock follows. The fish soon learns to anticipate shock whenever the frequency changes: It shows this anticipation by momentarily stopping its breathing.

[6] The experiments on goldfish were reported by Fay (1970) and Fay and MacKinnon (1969).

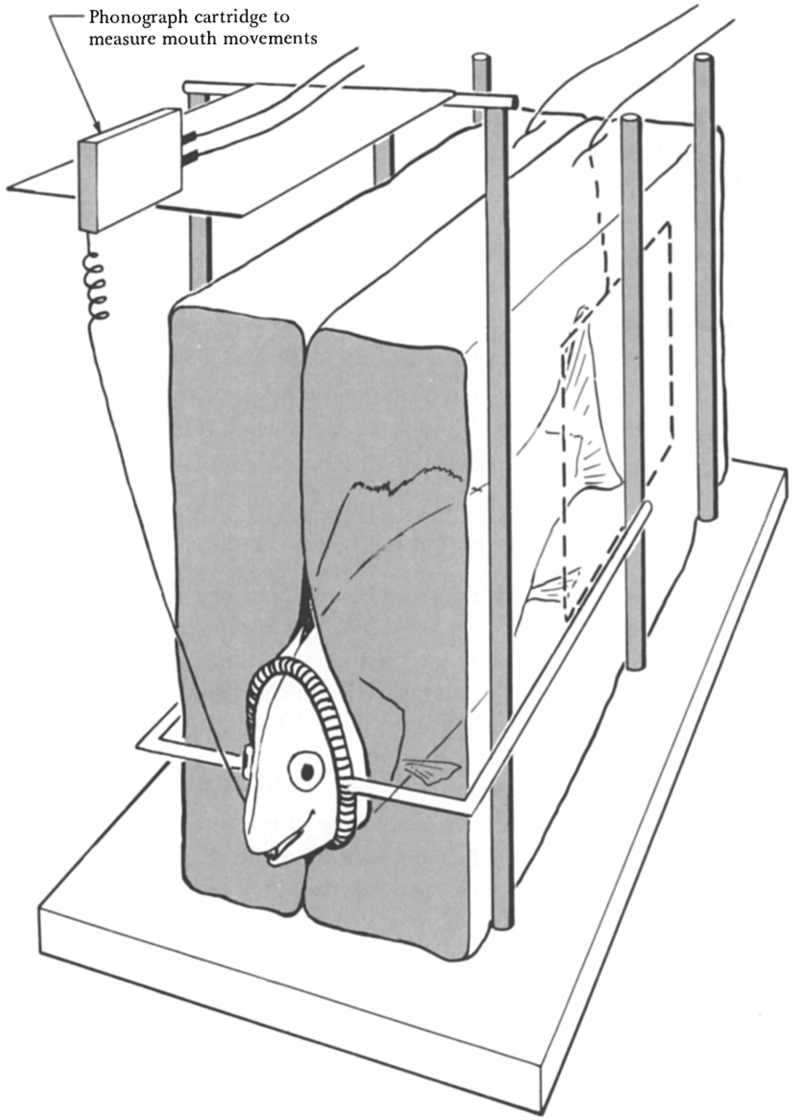


FIGURE 5-19

From Fay and MacKinnon (1969).

The data so obtained are shown in Figure 5-20. Note that the minimum frequency change that can be detected by the goldfish is about ten times larger than the minimum amount that can be detected by humans. Although the fish is much less sensitive in absolute terms, the way its sensitivity varies with frequency is similar to the corresponding human ability to discriminate. The ability of the goldfish to discriminate among frequencies disappears around 1000 to 2000 Hz, exactly what one would suspect from periodicity theory. At these levels, the nerves should no longer be able to fire in synchrony to the auditory signal.

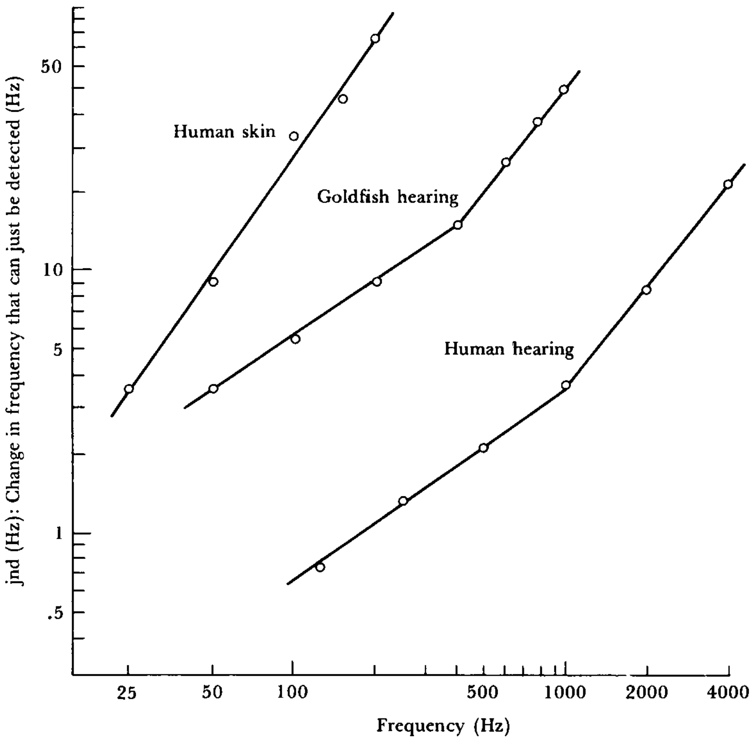


FIGURE 5-20

From Fay (1970).

Evidence against the periodicity pitch theory

Although the periodicity theory seems to fare pretty well in these masking experiments, there are two phenomena that give trouble to the periodicity account of pitch perception. One is that an individual nerve cannot respond more rapidly than about 300 to 400 times each second. How, then, can it be the basis for perceiving pitches corresponding to 4000-Hz signals? A second problem is the interesting anomaly of hearing called *diplacusis,* in which the same tone is perceived to have a different pitch with each ear. How could this be if the nerves simply fired at the rate at which the membrane was vibrating?

To answer the first criticism, periodicity theorists rely heavily on the volley principle (see page 143). Neurons fire in patterns, with a group of nerve fibers able to follow together a frequency that no single one of them could. If one nerve fires at a rate of only 300 Hz, a group of four neurons is capable of firing at a combined rate of 1200 Hz, given that everything is synchronized properly. Even so, there is no evidence that nerves, whether singly or in groups, can follow patterns of auditory frequencies greater than 2000 or perhaps 3000 Hz.

One of the strongest pieces of evidence against periodicity theory as a complete explanation is the anomaly of diplacusis, an ailment which lends credence to the theory that pitch is determined by the location of the maximum vibration on the membrane. Someone who suffers from a severe case of diplacusis will hear two different pitches when the same tone is played to each ear. Actually, everyone perceives some small differences in the pitch of a tone heard at the two ears, especially at high frequencies. The simplest explanation is that there is not a perfect match of positions along the basilar membrane. In fact, if you consider the precision of neural wiring that would be necessary to make each pair of locations on the two membranes correspond exactly, it is surprising that this phenomenon is not more prominent. It would be surprising even if the two membranes were exactly the same size, let alone matched neuron for neuron. Moreover, as one follows the neural processing up toward the brain, it is clear that there are many places where there might be a slight mismatch between the locations of neural fibers and the critical frequencies for which they are most sensitive.

The duplicity theory of pitch perception

We have just discussed two different ways by which the auditory system might determine the pitch of a sound. The two theories are called *place theory* and *periodicity theory:* the *place* on the basilar membrane where the traveling wave has its maximum bulge, and the rate and *periodicity* at which nerve fibers respond. Whenever there are two competing theories, both with good evidence for and against them, then some other path must be sought. Perhaps both theories are correct, but in limited ranges of operation. Alternatively, maybe there is some other way of looking at the phenomena that will encompass both these theories.

Consider how we said pitch information was extracted. The location at which the maximum vibration occurs on the basilar membrane, then, would appear to be a primary determinant of pitch, supplemented by information carried by the rate at which the fibers of the acoustic nerve respond. If the fibers located at the 1000-Hz location respond at a firing rate of 1000 Hz, everything is consistent. The perception is of a pitch given by a 1000-Hz tone. If the fibers at the 1000-Hz location respond in patterns of 100 Hz, then the perception is of a complex sound, with a fundamental pitch equal to that of a tone of 100 Hz, but with a harmonic structure in the 1000-Hz region. In this case, the firing rate helps determine the pitch: The location at which the membrane is responding determines the sound quality or *timbre.* Stimulation of a location along the basilar membrane is always accompanied by a firing rate appropriate to the location. Firing rates, however, are not always accompanied by stimulation of the appropriate membrane location. From our discussion, then, both place theory and periodicity theory may be correct: maybe there should be a *duplicity theory* that combines the two.

There are great difficulties with both the place and the periodicity (firing pattern) theories. Recent work has uncovered some basic problems but not yet provided a viable substitute. The most damning evidence against place theory is that exactly the same pitch sensation can be elicited from stimulating the high-frequency portion of the membrane as from the low-frequency portion (for example, in the "missing fundamental" experiment just described). The observer can easily distinguish the sounds, but the *pitch* of that perception seems unaffected by the location along the membrane that is stimulated.

Periodicity is not an adequate explanation either, for the waveform shape can be changed dramatically without changing the perceived pitch. Indeed, there is little change in the perception of the sound with gross changes in the waveform. In a concert hall, for example, the place where you sit determines the shape of the sound waveform. Echoes, differences in the path of high-frequency and low-frequency sounds, refraction patterns, and differences that result from sound amplification systems all cause quite different sound patterns at different seating locations. Moreover, you can cause dramatic changes yourself by cupping your hands over your ears and directing the open side of the cup toward the front of the hall or toward the back. This will significantly affect the higher-frequency components of the sounds that enter the ears, dramatically altering the shape of the sound waveforms. Nonetheless, although you will be able to perceive qualitative differences in the sound, the pitches that you perceive will be unaffected by choice of seat or by the position of your hands over your ears.

A few years ago, psychologists thought the problem of pitch encoding had been solved. The leading theory was a duplicity theory that combined both periodicity and place; low frequencies were analyzed by a temporal, periodicity analysis, whereas high frequencies were thought to be analyzed by place of excitation along the basilar membrane. Today, this view is held without reservations. The problem of determining pitch is now regarded as quite complicated, involving the integration of complex mechanisms within the brain. There is a considerable way to go before we can formulate a comprehensive theory to explain all the mechanisms that are involved.

Current theories tend to emphasize a centralized spectral (frequency) or temporal pattern-recognition system. That is, these theories emphasize the problem of analyzing the entire complex pattern of excitation along the neural pathway, rather than one single component such as place or rate of firing. In these theories, observers get some sort of global representation of the overall sound excitation from which they extract a generalized notion of pitch. These analyses are done centrally in the brain, not at the level of the basilar membrane or the hair cells. Thus, according to this view, both place and periodicity determine pitch: the entire pattern of excitation on the membrane is taken into account.

THE CRITICAL BAND

Suppose two pure tones are presented to a subject who is to judge the loudness of the resulting sound. As the two tones move farther apart in frequency (keeping their average frequency constant), the combined sound does not change loudness until a critical frequency separation is exceeded. Beyond this, the loudness of the tone pair increases with increased frequency separation.

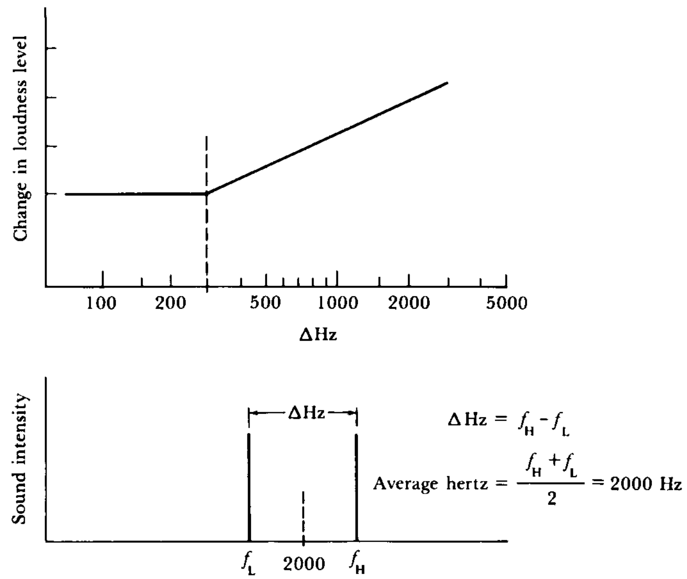


FIGURE 5-21

From Scharf (1970).

In a similar fashion, a sound containing components of all frequencies between some lower frequency f_L and a higher one f_H appears to have a constant loudness as the distance between f_L and f_H is increased, again until a critical value of separation is reached. From that point, the loudness of the sound increases as more and more frequencies are added.[7]

[7] Note that constant energy must be maintained in the sound (called *bandpass noise,* where the bandpass refers to the frequencies between f_L and f_H). To see how this is done, consider a simpler case in which a complex sound composed of separate tones is presented to a subject. If more tones are added, it is important for this task that constant total energy of the sound be maintained. Thus, if the number of tones is doubled, the energy of each tone presented must be halved to keep total energy constant. So it is with noise: The energy level at each frequency is kept proportional to $1/(f_H - f_L)$.

Consider a third example, a subject trying to detect a pure tone that is masked by bandpass noise centered around the frequency of the tone. As the distance between f_L and f_H is increased, detection becomes more and more difficult until a critical separation is reached. Beyond that point, detection is no longer affected by further increases in the width of the noise band.

All three of these examples indicate that within some critical region of frequency, sound energies are interacting with one another. As we move outside the critical region, sound energies no longer interact, although psychological attributes do add. The critical region is called the *critical band*. Its size depends upon its center frequency.

If we have two sounds that fall within the same critical band, their sound energies add. Thus, if we combined two sounds of equal energy within the same critical band, the effect is equivalent to doubling the *energy*. The resultant would be equal in loudness to a single tone that had *twice* the energy of either of the sounds. This is equivalent to a 3 dB increase in sound energy, and about a 20% increase in loudness. If the same two sounds were further separated in frequency so that they did not lie within the same critical band, then their *loudness* would add. If the two sounds had the same loudness, then when they were combined, the effect would be equivalent to doubling the *loudness*. The resultant would be equal in loudness to a single tone that had *ten* times the energy of either of the sounds. This is equivalent to a 10 dB increase in sound energy and a 100% increase in loudness.

If one looks at the pattern of excitation along the basilar membrane, it is surprisingly simple to find a correlate for the critical band similar

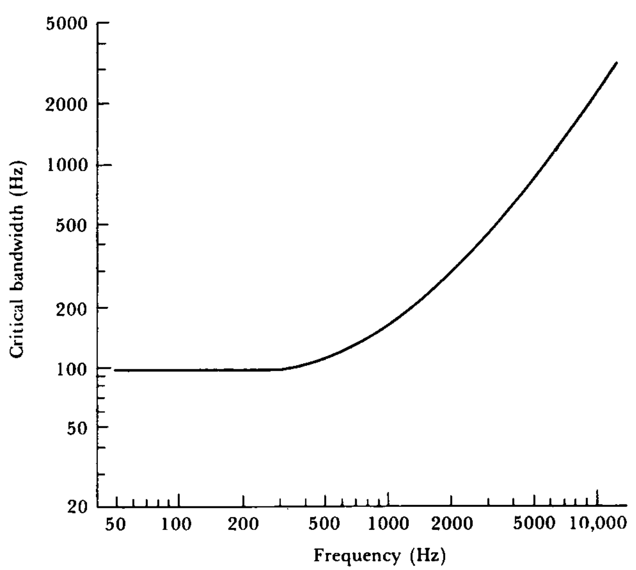
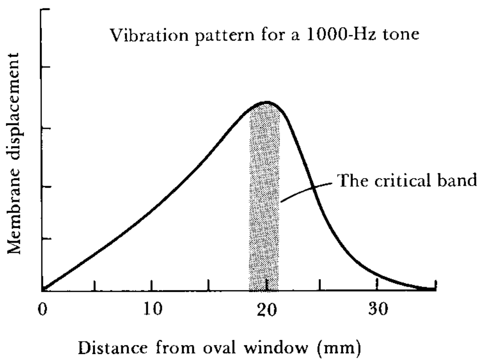


FIGURE 5-22

From Scharf (1970).

to the correlates found for the jnd and mel scales for pitch. The critical band looks as if it is simply caused by the vibration pattern. It is as if tones that vibrate nearby parts of the membrane interact differently from tones whose vibrations are further apart. Figure 5-23 shows the vibration pattern for a 1000-Hz tone and the part of the vibration pattern that corresponds to one critical band. It is interesting to note that although every part of the basilar membrane has some response to this tone, only a relatively narrow strip around the point of maximum vibration has the properties of interactions that are the characteristics of the critical band.

FIGURE 5-23



By comparing Figure 5-23 with the properties of the basilar membrane and the mel and jnd scales, the relationship among these factors can be nicely summarized:

- 1 mel is approximately 12 neurons, 0.23 jnd, and 0.009 critical bands;
- 1 jnd is approximately 52 neurons, 4.3 mels, and 0.04 critical bands;
- 1 critical band is approximately 1300 neurons, 108 mels, and 25 jnd's.

The critical band has many important properties. Beats between different tones appear to be noticeable only if the tones involved fall within the same critical band. The critical band has also been suggested as the mechanism responsible for the dissonance associated with some combinations of tones. Dissonance, it is argued, results from beats caused by two tones whose frequencies lie within a critical bandwidth of one another. Musical instruments produce complex tones, containing many harmonic frequencies. Two notes may be dissonant if any pair of their harmonics falls within the same critical band. The more audible these harmonics, the more dissonant the sound.

AUDITORY SPACE PERCEPTION

We have two ears, but we hear one acoustic world. With differences in information received by listening with two ears (*binaural*) rather than with one (*monaural*), we determine the locations of sound sources, an important factor both in adding to the enjoyment of our perceptions and in making acoustic messages more intelligible. It is difficult to appreciate the importance of sound localization because it is so seldom that we are without it: It is so common a phenomenon that we take it for granted.

With modern audio sets, however, the importance of sound localization is easy to demonstrate. Listen to a good, high-quality stereophonic recording over earphones.[8] Now, simply switch between monophonic and stereophonic modes. Listen to the difference. Stereophonic and quadrophonic reproduction not only allows the sounds to be perceived as originating from different locations in the imaginary space around you, but also gives a richer sensation of sound—one in which the various sounds are more distinct and easier to listen to.

Localization

The cues used to localize a sound source are the exact time and intensity at which the tones arrive at the two ears. Sounds arrive first at the ear closer to the source and with greater intensity. The head tends to cast an acoustic shadow between the source and the ear on the far side.

Figure 5-24

With some simple calculations, it is possible to determine the approximate maximum possible time delay between signals arriving at the two ears. The width of the human head is approximately 7 inches (18 cm). If a sound source is located directly to one side, the sound hits one ear directly but has to travel around the circumference of the head to get to the other ear. If the head is assumed to be a sphere with a radius of 3.5 inches, the extra path length is 3.5π, or 11 inches. Since sound travels at approximately 1100 feet sec^{-1} in air, it takes 76 microsec to travel an inch. For a sound to travel from one ear to the other takes around 840 microsec.

This time difference, of course, depends exactly on where the sound is located. When the sound is straight ahead, it reaches both ears at the same time. When the sound is 3° to the right, it arrives at the right ear 30 microsec before arriving at the left. This slight change—30 microsec of time difference—is detectable. It is all the change needed for an observer to detect a change in the location of the sound source. This is amazing performance, especially since the signals at the two ears must be compared with each other in order to localize the sound. The nervous system must be preserving information about the time at which a signal arrives at an ear within an accuracy of 30 microsec.

[8] You will get a less dramatic effect with speakers unless you sit so that one speaker is directly to your left and one directly to your right. The recording must be one that was recorded in the studio for stereophonic reproduction—old records or budget productions may not be very good for this. Most good rock groups or modern symphony recordings are excellent.

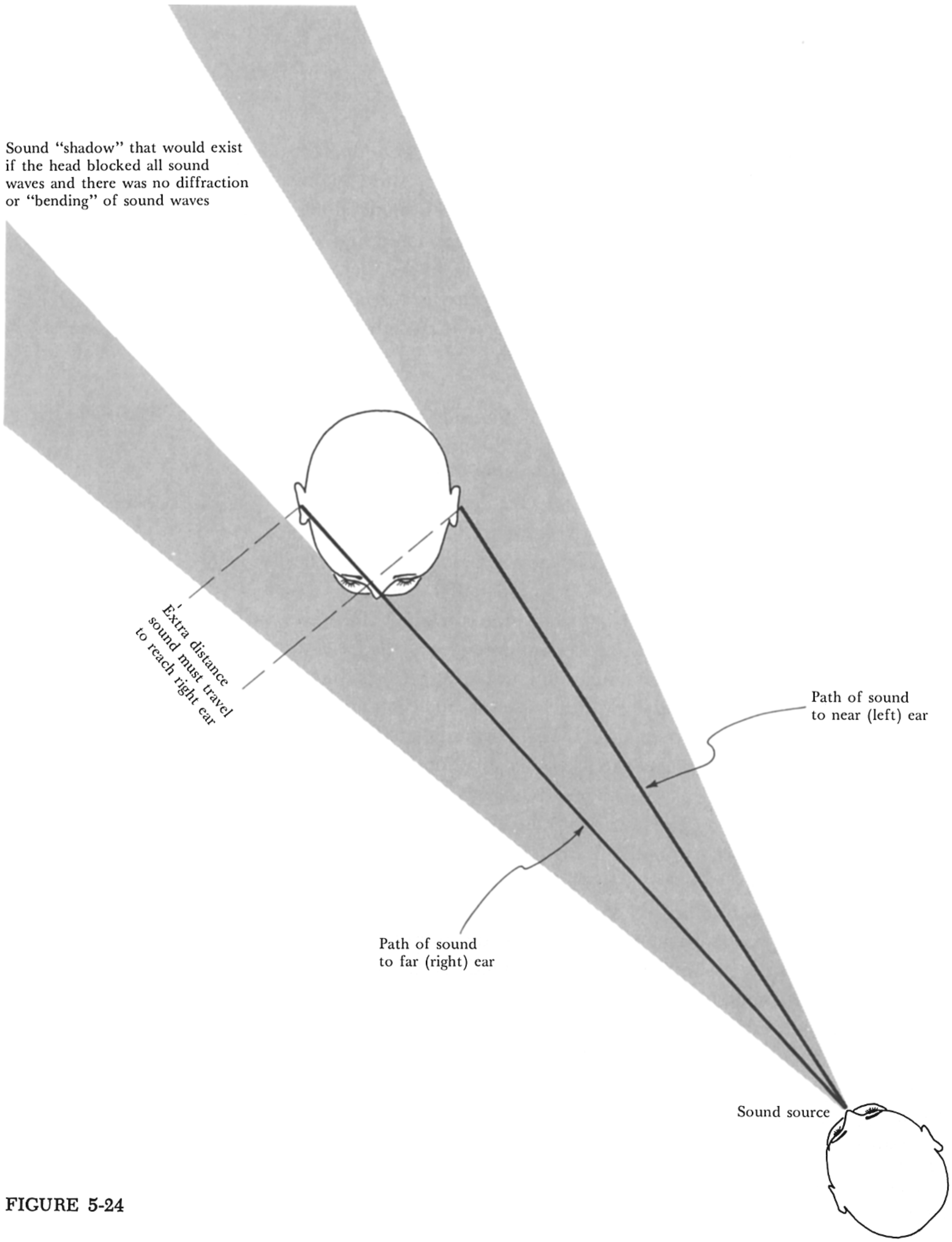
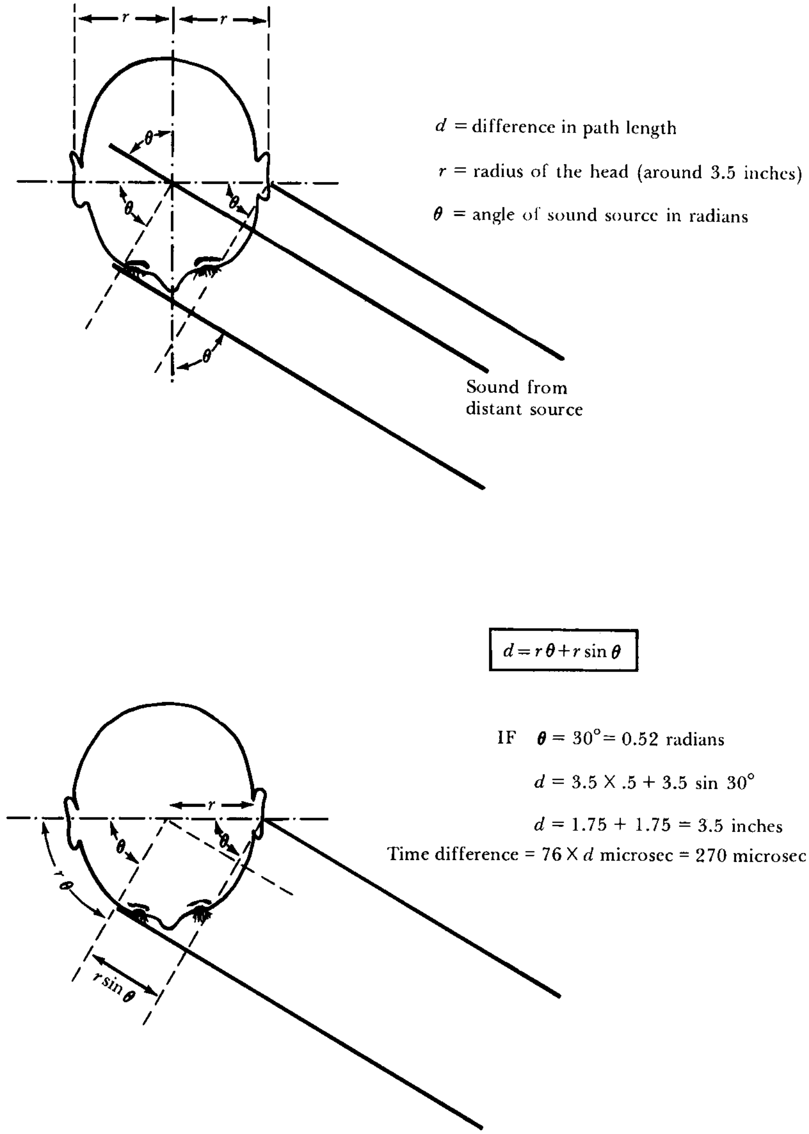

Sound "shadow" that would exist if the head blocked all sound waves and there was no diffraction or "bending" of sound waves
Extra distance sound must travel to reach right ear
Path of sound to near (left) ear
Path of sound to far (right) ear
Sound source

FIGURE 5-24



Approximate computation of binaural difference in path length for a distant sound source. **FIGURE 5-25**

The difference in the arrival time of a signal at the two ears results in a phase difference between the signals: One lags behind the other. For high-frequency signals, this time lag is ambiguous and cannot be used as a cue for localization. To see this, consider a 10,000-Hz signal. It completes a cycle of sound pressure variation every 100 microsec. When a 10,000-Hz signal is in front and to the right at an angle of 55°, the sound will arrive at the left ear some 450 microsec after it arrives at the right ear. The waveform heard at the right ear, then, is 4½ cycles ahead of the waveform at the left. But how can one tell whether two tones differ by

4½ cycles, 3½, 2½, 1½, or even ½ cycle? Alternatively, how can one tell if the source is 55° to the right or 40°, 27°, 17°, or 6°? There is no way. The longest time delay between the ears is around 840 microsec, and any sound frequency that takes less time than this to complete a cycle starts to give ambiguous location information. Time differences only provide good cues to location for sounds with frequency components less than about 1300 Hz.

Actually, localization is ambiguous even at low frequencies, since if the head is perfectly stationary, time difference alone cannot distinguish whether a sound comes from above or below, or even from front or rear. A sound in front and to one side has the same delay pattern as one in the rear in the same relative position. In real situations, these ambiguities can be removed by head movements, by visual cues, and by differences in sound quality caused by the way different frequencies are reflected and refracted by the head and outer ear.

A second cue to sound localization is the sound shadow cast by the head. With low-frequency sounds, the sound wave is diffracted and "bends" around the head, causing little or no shadow. But at high frequencies—when the wavelength is short compared with the dimensions of the head—diffraction does not take place to any significant degree. For example, a 100-Hz sound has a wavelength of 11 feet (3⅓ meters). Thus, it bends easily around the head. But a 10,000-Hz sound has a wavelength of only 0.11 feet (1.3 inches or 3.3 cm) so that it is reflected by the head, thus casting a shadow. With a source of sound at a 15° angle, the effects of the sound shadow can be measured:

Frequency	*Ratio of Sound Intensities at the Two Ears*
300 Hz	1 dB
1,100 Hz	4 dB
4,200 Hz	5 dB
10,000 Hz	6 dB
15,000 Hz	10 dB

Starting about 3000–4000 Hz, the intensity difference is great enough to be reliably discriminated, and thus provides a useful cue to localization.

Sound localization is carried out by a dual system: time differences for low frequencies and intensity differences for high frequencies. The switch between the two systems occurs in the frequency range of 1000–5000 Hz—the range of sound frequencies characterized by the largest amount of error in localization.

In determining the position of an object in space we use many different cues. So far, we have simply shown how time and intensity differences could determine the position of a sound source. But we can do more than that: we can tell the elevation (height) of the source, as well as the distance. The pinna of the ear (the skin and cartilage structure of the outer ear) plays a major role in our ability to localize sound sources. Sounds from the rear cast sound shadows because of the pinna, and the shadows affect high frequencies more than low ones. In addition, all those funny nooks and crannies of the pinna actually play an important role in determining the location of a sound, for they cause high-frequency components of the sound to bounce around a bit before entering the ear canal. This causes very slight "echoes" of the high-frequency portion of the sound wave, the exact sound depending upon the exact location of the sound source. Fill up those nooks and crannies with putty and the ability to tell the height and forward–backward direction of a sound decreases.

The distance of a sound source is judged by the reverberant nature of the sound. This cue is not reliable, but it still provides useful information, especially if the listener knows what the source would sound like from close up: then the differences in the sound can be used as a guide to its distance.

Importance of binaural listening

In addition to adding a spatial dimension to the perception of sound, binaural presentation also adds clarity. This is a consequence of three different mechanisms: localization, an apparent reduction in interference, and the minimization of masking.

Localization. Localization allows us to spread out in space many of the sounds that are heard. Suppose we are stuck with some boring people at a party. We can keep nodding and agreeing while actually attending to a neighboring conversation. Localization makes this possible. We can choose the frequency, intensity, or spatial location to which we want to listen.

When tape-recording a conversation, the result is often very difficult to understand. There are echoes and noises. The sounds of people coughing and moving about drown out the desired voice. In the real situation, we are not aware of these noises, even though they are present. Localization cues let us attend selectively only to the auditory signals of interests. A dramatic improvement in clarity comes by adding a second microphone—making a stereophonic recording. Suddenly, one can listen to **where** the voice is, tuning out the distractions. All that is needed are two microphones differentially sensitive to sound direction. They are set up properly if one is primarily sensitive to sounds coming from the right, while the other mostly picks up sounds from the left. When a person speaks from between the two microphones, he should be picked up equally on both of them. If he stands to one side, he will come through more on one than on the other.

The problems encountered in listening to a single-channel, monaural tape recorder dramatically illustrate the problems that must be encountered by individuals who are deaf in one ear. The difficulty is caused not so much by decreased sensitivity to sound, but by the reduced ability to localize sounds. If a hearing aid is used, where is the microphone to be placed? If the hearing aid microphone is worn in a shirt pocket, normal localization will not be possible. The microphone should be as close as possible to the ear. In fact, it might be best to use two hearing aids, one for each ear (even if one ear is normal), in order to recover the ability to localize sounds.

Masking level difference. The second way in which binaural reception improves clarity is by means of a phenomenon called *the masking level difference.* When trying to hear a weak voice mixed with noise presented to one ear, the addition of the same noise to the other ear will significantly improve the clarity. One ear has both signal and noise; the other ear has noise alone. One way of looking at this is to imagine the inputs to the two ears being subtracted from one another, causing the noise to be cancelled. Thus, putting the same signal and noise in both ears would do no good: Subtracting the inputs to the two ears leaves nothing. An alternative way to view this is to notice that the noise being presented to both ears is lateralized in the center of the head, whereas the signal is heard only at one ear. The difference in spatial location leads to the improvement in intelligibility.

Masking. There is a third way in which binaural reception can improve clarity. Imagine listening to an orchestra with a big bass drum banging away while a clarinet plays in the low registers. If this is recorded monaurally, the very low frequencies of the drum will mask the low frequencies of the clarinet. This results from the overlap of excitation along the basilar membrane. However, if the clarinet and drum are heard in opposite ears, there can be no interaction along the membrane: Masking should not occur. Obviously, in an actual concert, sounds from the clarinet and drum will get to both ears, but differently to each ear. When two sounds are localized as originating from two different locations, the masking effect of one upon the other is much reduced. This reduction of masking makes the sounds clearer and more distinct in a binaural than in a monaural recording.

The precedence effect

In theory, localization is performed simply by using differences in the sounds arriving at the two ears. Usually, however, the initial signal is immediately followed by numerous echoes. Once all the echoes are accounted for, even a simple click can be very complicated. First the click arrives at one ear, then the other, and then parts which have bounced off the walls and ceiling of the room start arriving at the ears.

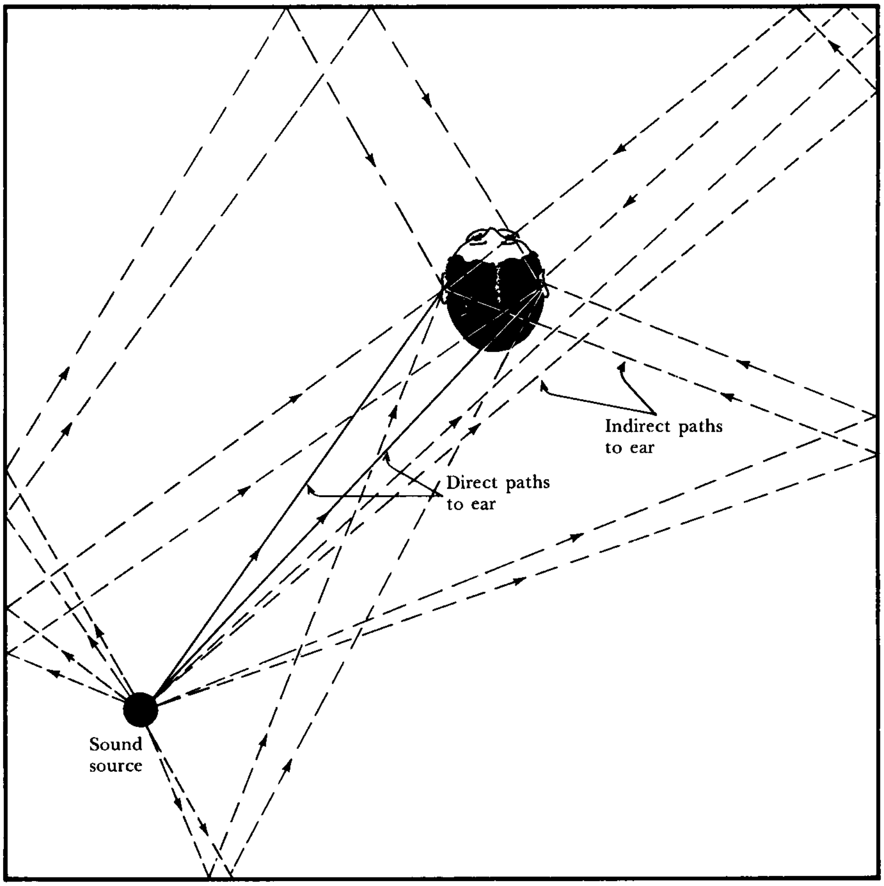


FIGURE 5-26

The two ears hear a rapid succession of sounds. How can one make use of all that to localize?

Fortunately, only the first sound to arrive appears to be used. This is called the *precedence effect*. It is not completely understood. Echoes play almost no part in the psychological interpretation of the sound. Not that they are not heard. If we record various sounds, some followed by echoes and some not, it is easy to tell the difference between them. Thus, the sound information is heard, but fortunately, it is ignored by the mechanism responsible for localization.

Recordings

Binaural sound. To get a recording of an event that really sounds as if you were sitting in the auditorium, it is necessary to make a *binaural* recording. To do this properly requires that a dummy of a head be placed in a seat in the auditorium, with microphones in each of the dummy's ears. When played back over earphones, binaural recording gives beautiful fidelity.

Stereophonic sound. Binaural recording is very different from the *stereophonic* recordings that are used for most records and tapes. In

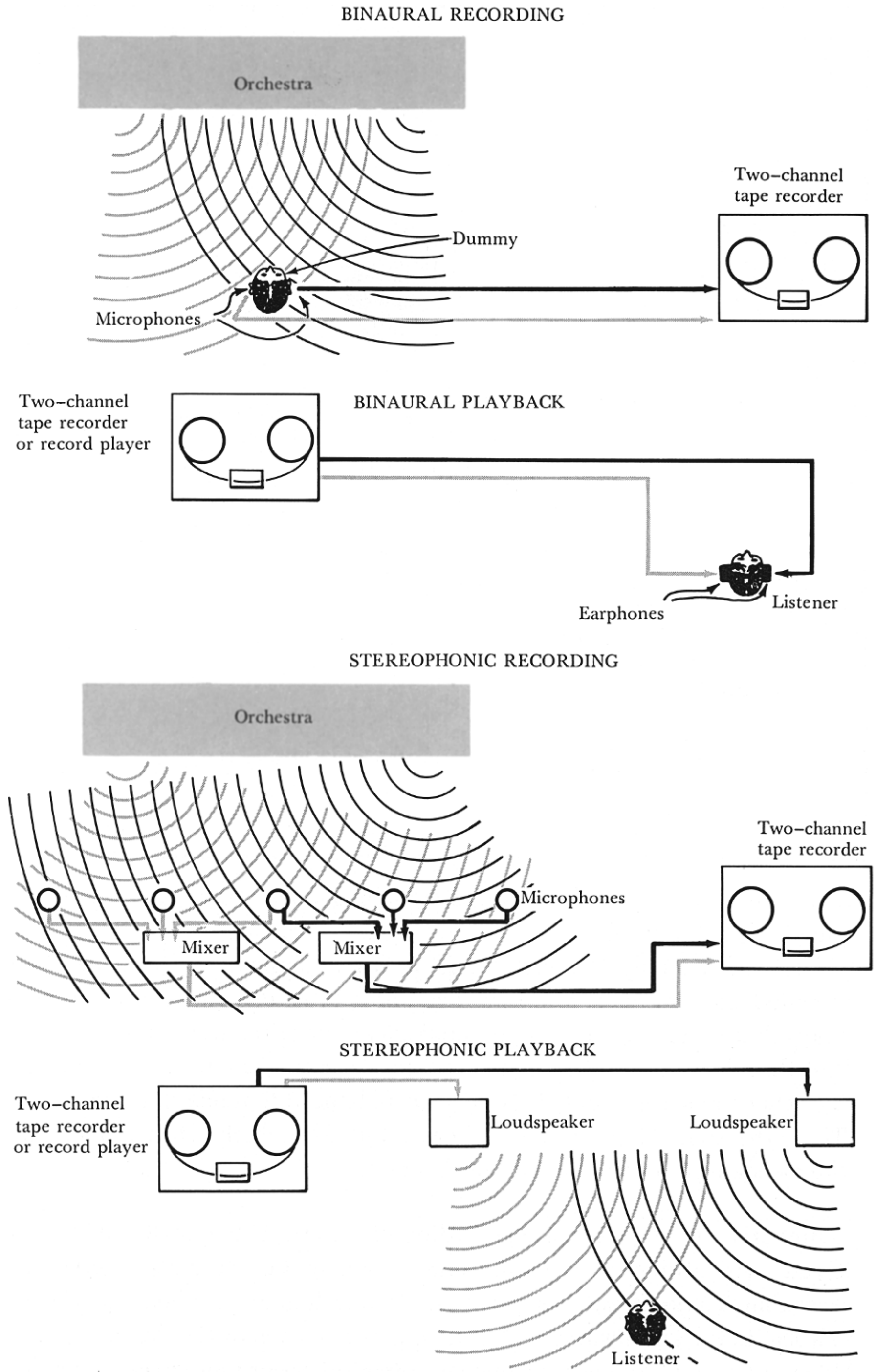
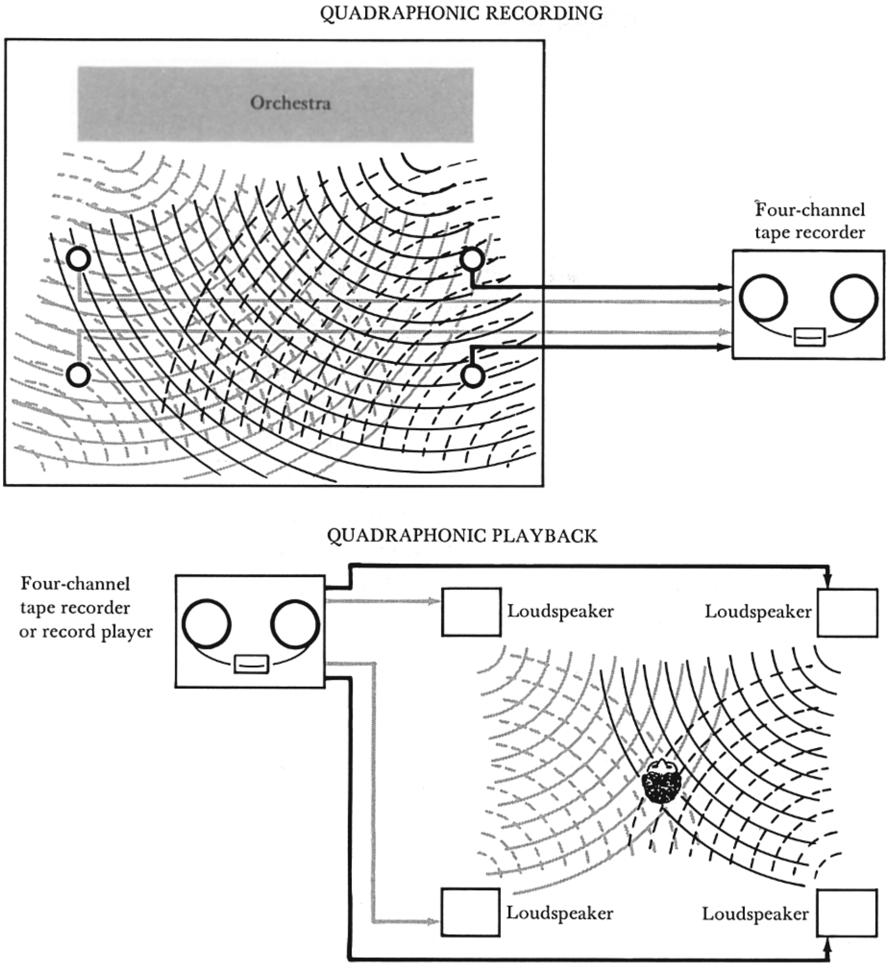


FIGURE 5-27

stereophonic recording, the object is to try to slice across the wavefront as it crosses some point in the auditorium and reproduce it in the listener's home. There is no way that this can actually be done accurately with only



two microphones and two speakers. For this reason, four-channel sound is now being tested for home use. Ideally, with two speakers there should be two microphones spaced the same distance apart as the speakers will be. This is not very good, since it demands too much in the way of control of the playback conditions. As a result, recording engineers have learned to combine many microphones, mostly by trial and error. Pleasing results can be obtained in this manner, so that is the way recordings are currently being done. Generally, in order to make a really good recording many microphones must be scattered over the auditorium: The recording engineers determine how to combine them into two channels to give the proper effect. Psychological acoustics does not help much here. The best way to combine channels probably differs for every auditorium, for different numbers of people in the audience, and even for the way the audience is dressed.

Quadraphonic sound. As Figure 5-27 shows, the sound that we normally hear includes reflections. The feeling of spaciousness of a symphony hall comes about, in part, as a result of these reflections. Thus, stereophonic sound fails to capture all the qualities of the original sound experience, in part because the two speakers are both placed in front of the listener. This limitation is not a problem with binaural sound, because here the listener experiences exactly the same sounds as a person sitting in the auditorium. The problem with binaural sound is that it is designed for one particular position of the head: if the listener's head moves, the earphones move along with it, and so the original sound sources always appear to come from in front of the head. This gives an uncanny power to the listener. Each head movement makes it seem like the entire orchestra and room also move to keep directly in front of the head.

One technique for enriching the quality of the sound experience without the limitations presented by binaural sound recordings is to use four speakers, two in front and two to the rear of the listener. Four speakers offer a distinct improvement in the faithfulness of the situation, although it too must suffer from being only a poor approximation to the original event. With four speakers, the effect of the ears' pinnae becomes important in distinguishing sounds from the front of the head from sounds behind. The use of four speakers opens up new vistas for auditory experiences. Some composers have experimented with musical presentations aimed explicitly at being presented over four (or more) loudspeakers, with the spatial dimension of the musical experience being varied directly, in much the same way as loudness, pitch, or sound quality. (See the discussion by Reynolds, 1975, pp. 117–125.)

REVIEW OF TERMS AND CONCEPTS

In this chapter, the following terms and concepts that we consider to be important have been used. Please look them over. If you are not able to give a short explanation of any of them, you should go back and review the appropriate sections of the chapter.

Terms and concepts you should know

The differences between:
- pitch and frequency (mels and hertz)
- loudness and intensity (sones and decibels)

Equiloudness contours
Loudness compensation
Loudness (sones)
Masking
- what it is
- how it is measured
- how it is explained

jnd
Pitch
- mels

musical scale
periodicity pitch
missing fundamental
beats
Place theory
Periodicity theory
Duplicity theory
Critical band
Localization
by arrival times
by intensity difference
ear shadow and masking

SUGGESTED READINGS

Perhaps the easiest place to start is with Volume IV of the *Handbook of perception: Hearing* (Carterette & Friedman, 1976). The two books *Foundations of modern auditory theory* edited by Tobias (1970, 1972) include several important chapters, especially on the measurement of loudness, theories of pitch, masking, the critical band, and localization.

We highly recommend the book *Mind models: New forms of musical experience* by the composer Roger Reynolds (1975) for anyone interested in contemporary uses of sound in musical experiences. Roederer's (1975) book *Introduction to the physics and psychophysics of music* is a more technical introduction to the relationship between the concepts discussed in this book and musical perception.

The chapter by Zwislocki (1965) in the *Handbook of mathematical psychology* (Volume III) is excellent, although a bit advanced. Do not be too dismayed by all the equations in the first part of the chapter. We have found that introductory students can get much meat out of this chapter if they simply skim quickly over the first few sections. The latter part of the chapter is especially valuable in discussing the relationships between the anatomy and physiology of the ear and the perception of loudness, pitch, and masking. The article by Wightman and Green (1974) is a good introduction to pitch.

The book by Kryter (1970) covers in detail *The effect of noise on man* (which is also the title). Unfortunately, although the material covered by the book is of utmost importance, the level of writing makes it very difficult for the reader.

A large amount of the fundamental advances in our knowledge of pitch was provided by the Dutch, especially the efforts of Schouten. The paper by Wightman and Green (1974) serves as an introduction to this literature, and the thesis by Plomp (1966) is especially important for an analysis of consonance and dissonance.