1 The Ear and How It Works

Max Mathews

1.1 Structure of the Ear

In this chapter we will introduce the structure and function of the ear. We will also talk about some of the other physical aspects of how sound gets into our ears, such as the effects that the head and shoulders have on what we hear. The outer ear or pinna (plural pinnae) leads to the middle ear's auditory canal or meatus. The auditory canal terminates at the eardrum, or tympanic membrane. Beyond the eardrum is the inner ear, which contains the hidden parts of the ear, encased in bone. The complicated structure is hard to see or to understand in photomicrographs of actual dissected specimens. We will rely on schematized drawings that depict the shapes and relations of the various parts of the ear. Even here we are in difficulty, for few diagrams give typical dimensions.

The schematized drawing of the ear in figure 1.1 shows the semicircular canals, three liquid-filled passages that are associated with equilibrium rather than hearing. They tell us about the orientation of the head, cause us to get dizzy when they are malfunctioning, and cause some of us to get seasick when the head, body, and eyes undergo motional disturbances. The three little bones of the air-filled middle ear, which are attached to the eardrum, excite vibrations in the cochlea, the liquid-filled inner ear. It is in the cochlea that the vibrations of sound are converted into nerve impulses that travel along the auditory nerve, the *eighth nerve*, toward the brain. The purpose of the auditory canal is to guide sound waves to the eardrum. The pinna acts as a collector of sound from the outside world and also as a directional filter, as will be discussed later. Other body structures outside of the ear also affect the sound that reaches the inner ear.

1.2 Acoustic Shadows and Filters

The intensity of a sound wave in the auditory canal is proportional to the intensity of the sound wave that approaches the listener. It also depends on the direction from which the sound wave comes.

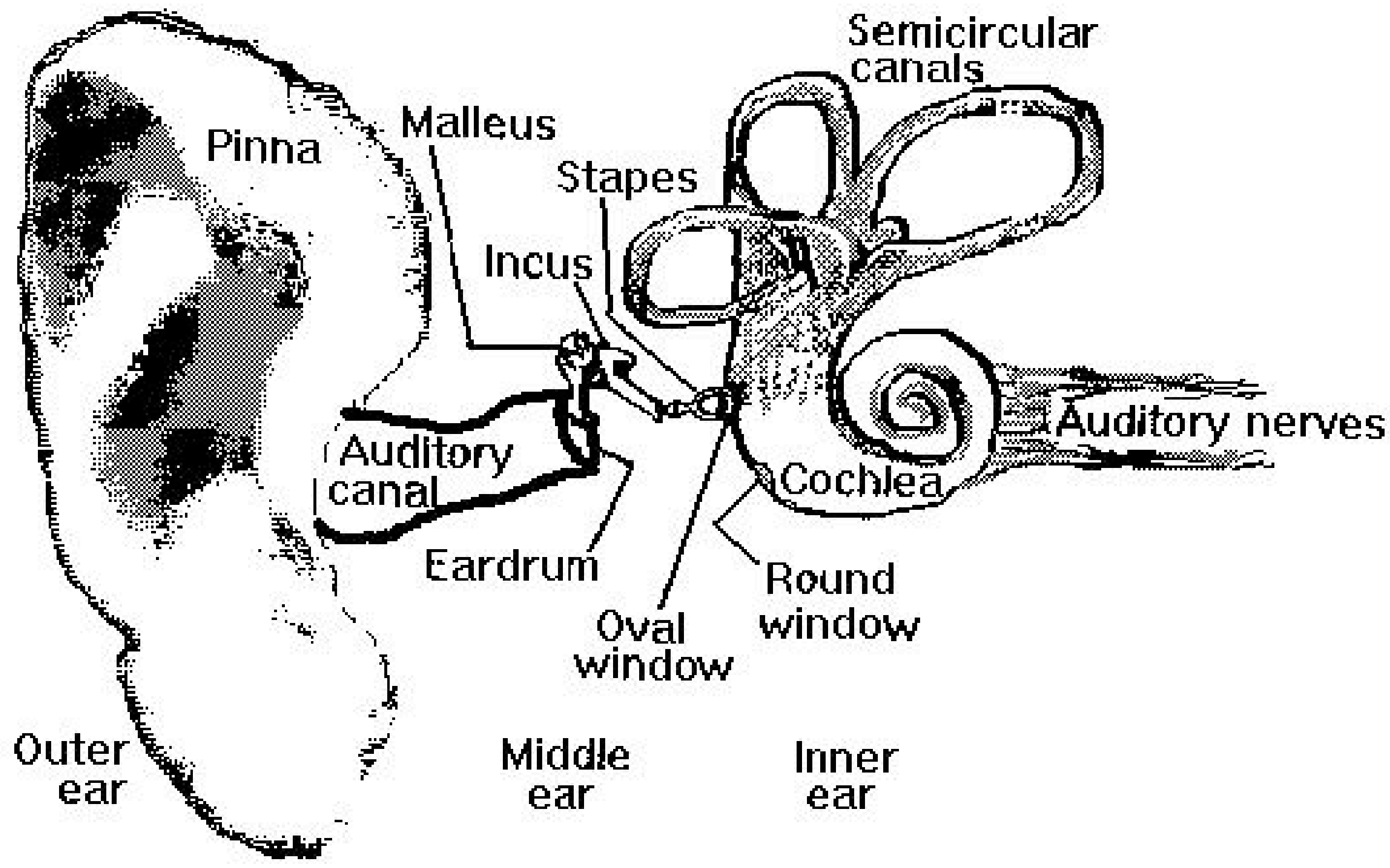

Figure 1.1 The outer, middle, and inner ears (not to scale).

The shoulders can reflect the sound wave toward the pinnae. The head can cast an acoustical shadow. The "guiding" effect of the convolutions of the pinna increases with increasing frequency, and at high frequencies it rises and falls with frequency. The effects of the pinna, the head, and the shoulders become significant only when the wavelength of the sound wave is comparable with the dimensions of the structures. Each shoulder is about 6 inches wide. The head is approximately an 8-inch sphere. The height of the outer ear or pinna is around 2 inches.

Sound moves through the air as a travelling wave with a velocity of 1128 feet (344 meters) per second. The wavelengths of sinusoidal sound waves of various frequencies and musical notes are shown in table 1.1. Middle C on the piano keyboard (traditionally located closest to the first letter of the brand name above the keyboard) is C4 and has a frequency of 261.6 hertz (Hz; cycles per second). Each doubling or halving of frequency is called an "octave," and causes the index number to go up or down by 1. Thus the C above middle C is C5, and has a frequency of 523.2 Hz.

From the tabulated values we would expect an appreciable influence from the shoulders and the head above the pitch of A6. We would expect pinna effects to be most important above a pitch of A8 (although no musical instrument normally produces a fundamental pitch this high). However, in most musical sounds much of the overall energy is contained in harmonics (integer multiples) of the fundamental frequency. Thus, reflections, shadowing, and guidance effects will occur at frequencies two octaves, three octaves, or even further above the actual pitch frequency.

Table 1.1 Musical pitch, frequency, and wavelengths

MUSICAL PITCH	FREQUENCY (HZ)	WAVELENGTH (IN.)
A2	110	123
A3	220	62
A4	440	31
A5	880	15
A6	1,760	7.7
A7	3,520	3.8
A8	7,040	1.9

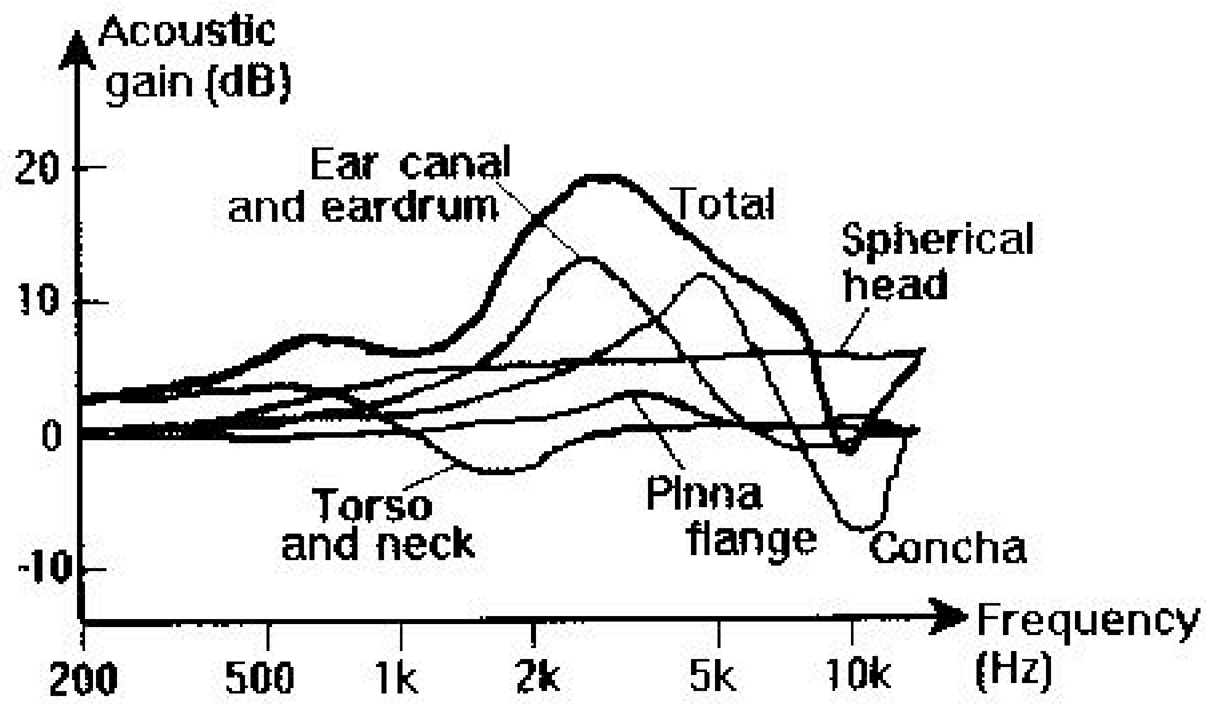

Figure 1.2 The acoustic effects of various body structures as a function of frequency. Sound is from 45 degrees off center in front of listener. (Modified from E. H. G. Shaw, 1974, *Handbook of Sensory Physiology*, vol. 5/1. Berlin: Springer, fig. 11.)

Computations have been made concerning variation with frequency of the ratio (measured in decibels, dB) of sound pressure at the eardrum to sound pressure in the wave approaching the listener. Figure 1.2 shows the outcome of such a calculation for a wave approaching from 45 degrees off center. The curve labeled Total shows the total variation with frequency, plotted in dB as a function of frequency. The other curves show the parts of the effect attributable to various structures of the body and outer ear. The concha and the pinna flange are separate parts of the pinna. We may note that many of the interesting effects happen between 2000 and 5000 Hz, with the total peak at around 2500 Hz.

The influence of the head in shadowing the ear from the sound source is important in judging left or right, particularly for high-frequency sounds. The disparity in time of arrival at the two ears is more important. Changes in the disparity by 20 or 30 microseconds (millionths of a second) can produce noticeable changes in our judgement of the angular position of a sound source.

1.3 The Pinnae and Auditory Canal

The frequency-dependant effect of the pinnae on the intensity of sound in the auditory canal is important in judging whether a sound comes from ahead or behind; it is crucial in judging the height of a source in the median plane (a vertical plane that divides the head into left and right halves). If you fold down the tops of your ears, you will notice that it becomes difficult to discern the position of a sound source.

The externalization of a sound source (the sense that the sound comes from some point outside the head) is ordinarily absent in listening through headphones, which give us only a sense of left and right. Some degree of externalization can be obtained with headphones by recording the two stereo channels from microphones located in the two auditory canals, or even in the auditory canals of a dummy head with typical pinnae.

The auditory canal is a little over an inch (2.54 cm) long and a quarter of an inch (0.6 cm) in diameter, and leads from the pinna to the eardrum. If we regard the auditory canal as a quarter-wave resonator (this will be discussed more in later chapters), the ratio of sound pressure at the eardrum to sound pressure at the outer opening of the canal will be greatest at a frequency around 3000 Hz. From this alone, we can deduce that the maximum sensitivity of the ear to sound should be around 3000 Hz.

1.4 Inside the Ear

The eardrum vibrates in response to sound. For us to hear this vibration, it must be transferred from the air-filled outer and middle ears to the fluid-filled cochlea, which is "stiffer" (has a higher impedance) than air. Transfer of vibration is effected through the leverage of a chain of three little bones or *ossicles,* the *malleus* (hammer), the *incus* (anvil), and the *stapes* (stirrup).

The chain of little bones can be stiffened and the transfer of vibration decreased by the contraction of the *stapedius* muscle. This provides a gain control that gives some protection to the inner ear, because loud sounds cause a contraction of the muscle and a reduction in the sound transmitted to the cochlea. This *stapedius reflex* gives some protection against steady sounds, but no protection against very intense, very brief sounds, such as a gunshot.

Figure 1.1 shows the cochlea as a coiled, snail-shell-like structure. Its actual size is about that of a pencil eraser. Vibration is transmitted into the cochlea by the stapes driving the flexible membrane that covers the oval window. Below the oval window is the round window, which also is covered with a flexible membrane; the function of the round window will be discussed later. The auditory nerve (*eighth nerve*, or *spiral ganglion*) comes from the cochlea, carrying *afferent* nerve impulses toward the brain, as well as some *efferent* nerve impulses from the brain into the cochlea. The function of the cochlea is to convert the vibration of sound into nerve impulses in the auditory nerve. The details of the way it carries out this function can be modified by pulses reaching it from the eighth nerve. Let us now consider the structure and function of the cochlea.

The cochlea is a long, narrow, three-and-a-half-turn coiled passage surrounded by bone. The passage is about 35 millimeters (mm) (1⅜″) long. The diameter of the passage (around 2 mm, or ⁵⁄₆₄″) varies somewhat with distance from the basal, oval-window end, being least at the apical end. Figure 1.3 shows a cross section at some place along the tube of the cochlea. From this figure we see that the cochlear passage is divided into three parts:

1. The *scala tympani* stretches from the oval window at its basal end to the apical end. The oval window is driven by the stapes at the basal end. The apical end is the end at the small termination of the coiled cochlea.
2. The *scala vestibula* stretches from the round window at its basal end to the apical end.
3. These two scalae are connected at their apical end by a passage called the *helicotrema* (not shown in figure 1.3).

The *scala media* lies between the other two scalae, and is closed at both the basal and the apical ends.

All three scalae are filled with fluid. The fluid in the scala media is different from that in the other two, and the difference in composition makes the scala media fluid (*endolymph*) about 80 millivolts positive with respect to the *perilymph* that fills the other two

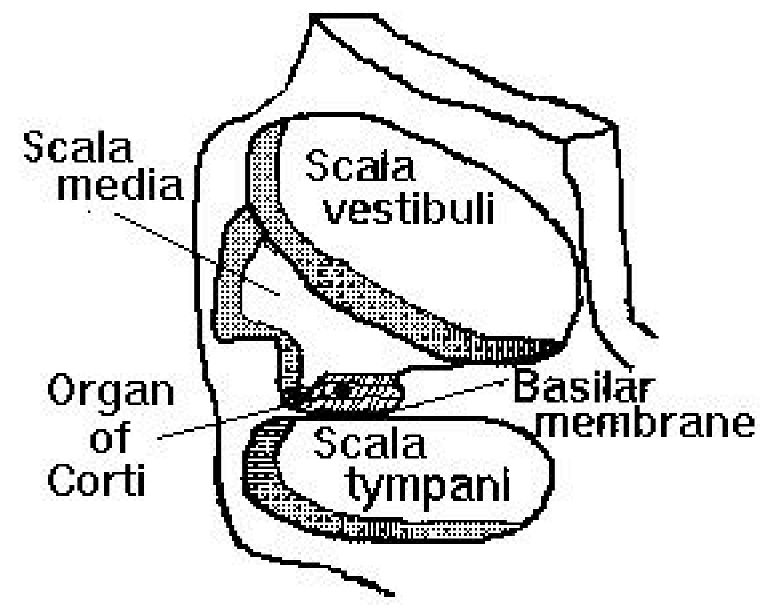

Figure 1.3 Cross section of the cochlear ducts.

scalae, and also with respect to the bony tissue of the cochlea. This potential difference plays a part in the excitation of nerve pulses.

The actual production of nerve pulses takes place in the *organ of Corti.* At the bottom of the organ of Corti is the *basilar membrane,* which separates the scala media from the scala tympani. The basilar membrane vibrates in response to sound, and that vibration causes vibration in some of the 3500 inner hair cells with respect to the *tectorial membrane* that lies on top of them. This relative motion produces nerve pulses that are transmitted toward the brain by the eighth nerve. There are also three rows of outer hair cells whose functions are different. They respond to nerve pulses coming toward them *from* the brain via the eighth nerve.

1.5 Details of the Cochlea

To understand hearing, it is important to understand the way in which the basilar membrane responds to the vibrations of the stapes pushing the oval window in and out. This mechanical aspect of the cochlea is made more obvious by "unrolling" the cochlea and its basilar membrane into a tapered tube, as shown in figure 1.4. The basilar membrane divides the tapered cochlea into two passages that are joined at the right end by the helicotrema. The upper passage, which starts at the oval window at the left or basal end, represents the scala tympani. The lower passage, which starts at the round window at the left, represents the scala vestibuli. The scala media, closed at both ends, is not shown.

We may regard the fluid in the cochlea as incompressible. In figure 1.4 the oval window is shown at the top, and the round window at the bottom, of the basal end of the cochlea. If the oval window is pushed in slowly, the round window will bulge out slowly as fluid

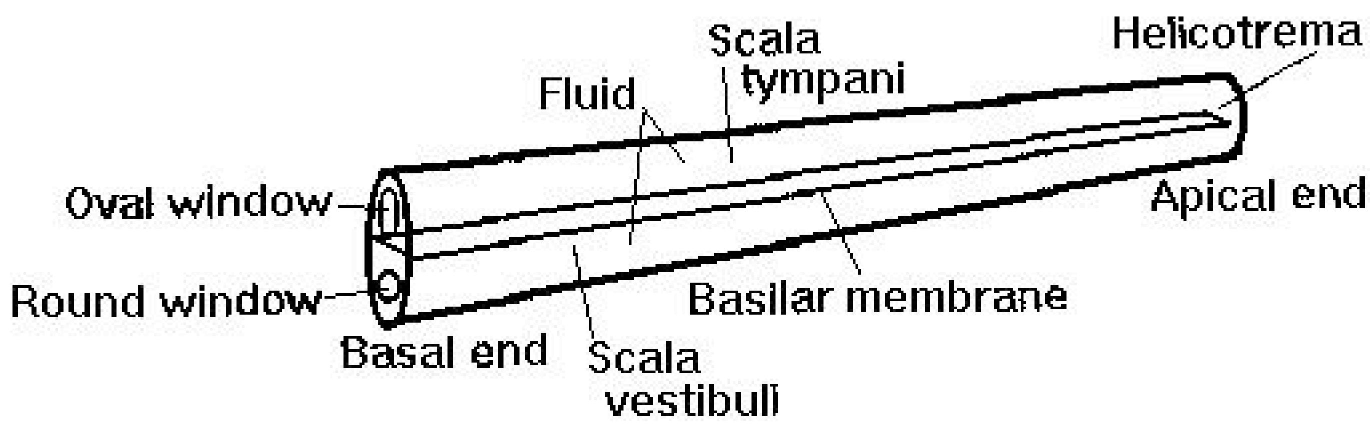

Figure 1.4 An unrolled view of the cochlea. Fluid volume is constant. If the stapes pushes the oval window in, the fluid pushes the round window out.

passes through the helicotrema slowly, so that the total volume of the fluid remains constant.

What happens if we push the oval window in quickly? The springy basilar membrane is pushed down, the oval window is pushed out, and constant volume is maintained without motion of fluid through the helicotrema. In fact, when the oval window is pushed in suddenly, the deflection of the basilar membrane is a wave that travels from the oval window toward the helicotrema. As in an ocean wave or a sound wave, there is no actual flow of liquid from the basal to the apical end. Rather, deflection of the basilar membrane at one point causes a subsequent deflection of the adjoining part of the basilar membrane, and so on.

Suppose we excite the oval window with a steady sine wave. Figure 1.5 shows deflection vs. distance at successive times, moving rightward. We see that as the wave on the basilar membrane travels from left to right, it slowly increases in amplitude, peaks, and then rapidly decreases in amplitude. In essence, the wave of this particular frequency can't travel beyond a certain "cutoff point" on the cochlea. As it approaches this place, the wave slows, the amplitude increases, and the power of the wave is absorbed. A wave of that frequency can't travel beyond the cutoff position. This cutoff phenomenon is important because a wave of high frequency can't mask (or hide) a wave of low frequency, while a wave of low frequency can mask a wave of higher frequency.

The dashed lines in figure 1.5 designate the envelope of the wave as a function of distance along the cochlea, the highest (or lowest) amplitude as the wave passes. Figure 1.6 shows the envelopes of waves along the cochlea for various frequencies. We see that the basilar membrane acts as a filter in which the maximum response to a sinusoidal sound wave moves further to the right as the frequency of the wave is decreased. Thus, the basilar membrane excites different hair cells for sufficiently differing frequencies.

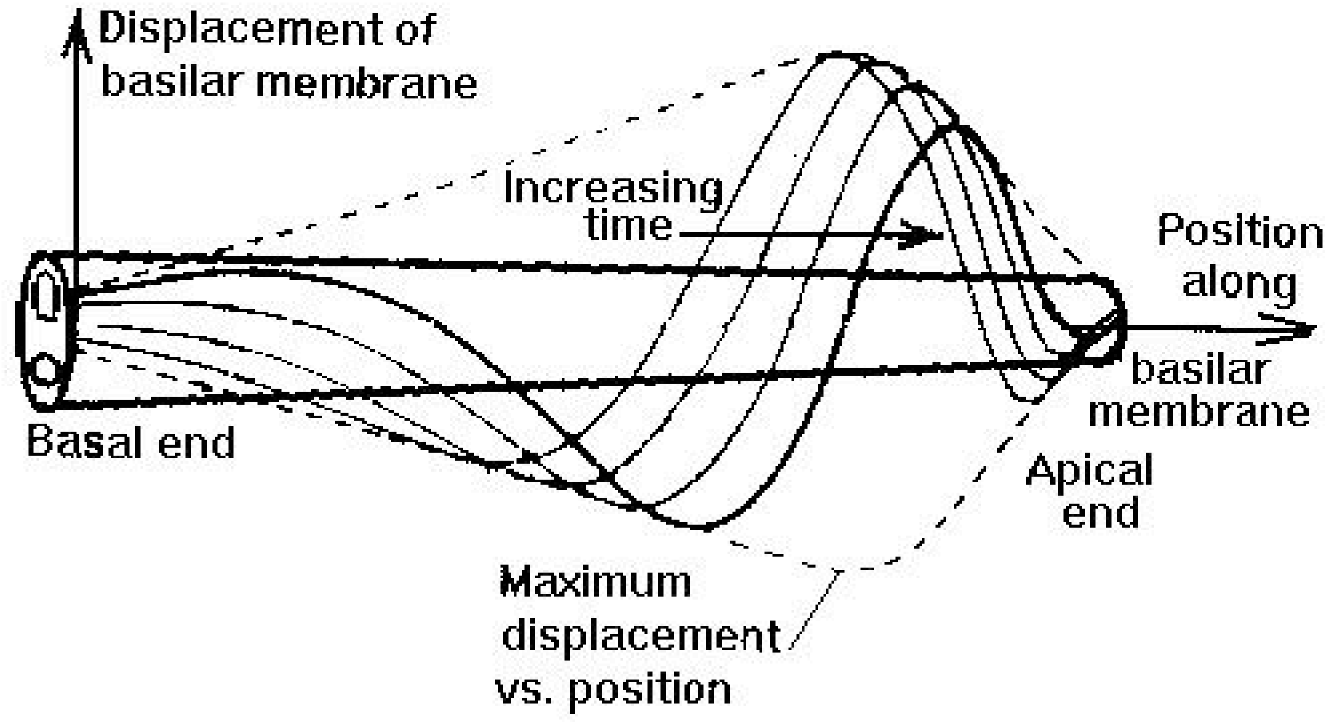

Figure 1.5 Progressive basilar membrane deflection vs. distance for a 200 Hz wave. The dashed curves show the envelope, the maximum upward and downward deflection, as a function of position along the basilar membrane. (Modified from G. von Békésy, 1960, *Experiments in Hearing*. New York: Wiley, fig. 12–17.)

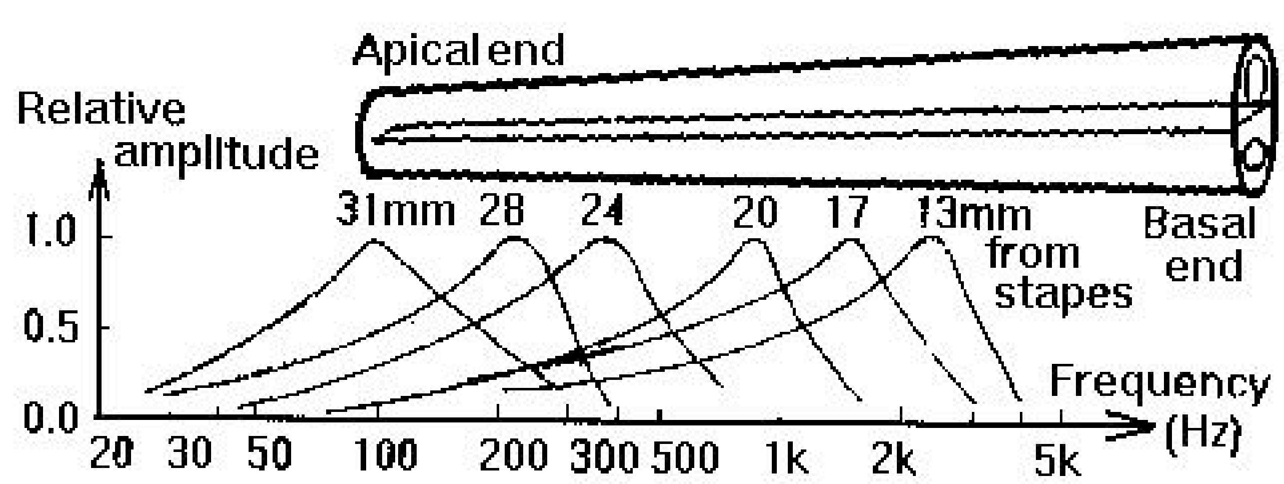

Figure 1.6 The envelope or maximum deflection as a function of distance along the basilar membrane for sinusoidal excitation at the oval window of various frequencies. (Modified from G. von Békésy, *Experiments in Hearing*. New York: Wiley, 1960, fig. 11–49.)

What happens when a sound consists of two frequencies that aren't widely separated, so that some hair cells are excited by both sinusoids? We hear beats. If the two frequencies are a few cycles apart, what we hear is like a single sine wave of rising and falling intensity. As we increase the frequency difference, we hear a vibrating, somewhat harsh sound. If we make the frequency difference large enough, we hear the two sinusoidal sounds clearly and separately.

Thus, we can explore the frequency resolution of the cochlea by listening to pairs of sine waves and varying their frequency separation. Or, at a given frequency, we can add to a strong sine wave a weaker sine wave of different frequency whose intensity is some number of dB less than that of the strong sinusoidal signal. For each frequency separation we can ask, "How much weaker [in dB] must we make the weaker signal in order that it be masked [not heard] in the presence of the stronger signal?" In this way we can obtain *masking curves* for different frequency separations. A *critical band-*

width is the bandwidth within which two sinusoidal signals interact, or mask one another. It should be possible to relate the masking curves to curves of response vs. distance, such as those of figure 1.6. Measurements of critical bandwidth indicate that the curves of figure 1.6 aren't narrow enough. The curves of figure 1.6 are based on early work of Békésy (work that won him the Nobel Prize). Békésy measured cochlea extracted from corpses. More recent data from experiments on living animals show that the curves in living creatures are much sharper, or more narrow-band, than Békésy's data indicate.

Much effort has been devoted to mathematical models of waves traveling along the basilar membrane. Such models are intended to take into account the stiffness of the membrane, which decreases from the basal end to the apical end, the mass of the fluid in which the membrane is immersed, and losses along the membrane. No model fits the data satisfactorily at any signal level. Moreover, a model that fits a dB vs. frequency curve fairly well at one sound level fits worse at other sound levels.

It is now believed that at low levels the outer hair cells act like little muscles and actually add energy to the traveling wave to make up for mechanical losses in the basilar membrane. Indeed, in some forms of tinnitus (ringing in the ears) the ringing can be picked up with a sensitive microphone placed near the ear—in such a case the loss has been more than compensated for, and some part of the basilar membrane oscillates.

The fraction of energy added by the outer hair cells is thought to be controlled by nerve signals carried to the outer hair cells by the auditory nerve. In effect, there is a gain control that reduces the response of the ear to sounds that send strong signals up the auditory nerve toward the brain. This gain control is much faster-acting than the stapedius reflex.

1.6 Concluding Remarks About the Ear

While not all details of the precise functioning of the ear have been fully worked out, several things are clear.

The basilar membrane of the cochlea performs a frequency analysis; a given place along the basilar membrane responds most strongly to a sinusoidal vibration of a particular frequency. While a sound of one frequency can mask a sound of a nearby frequency, it cannot mask a sound of widely separated frequency. Lower-frequency sounds can mask sounds of higher frequency, but higher-frequency sounds cannot mask sounds of lower frequency.

Sounds widely separated in frequency are heard separately. Sounds separated in frequency by less than a critical bandwidth interact to produce beats or harshness.

The ear provides two sorts of gain control so that we can hear sounds over a range of intensity of around 120 dB. One gain control is the slow-acting stapedius reflex. The other is provided through the three rows of outer hair cells.

References

Békésy, G. von. (1960). *Experiments in Hearing*. New York: Wiley.

Pickles, James O. (1988). *An Introduction to the Physiology of Hearing*. Second edition. San Diego: Academic Press.

Pierce, J. R. (1992). *The Science of Musical Sound*. Revised edition. New York: W. H. Freeman.

8 Hearing in Time and Space

John Pierce

8.1 Hearing the World

Sight tends to dominate hearing. When we watch home movies or TV, we hear the voices as coming from the characters on the screen, regardless of where the loudspeakers are located. But without visual clues we do sense the direction and even the distance of sound sources.

Imagine stepping down from the porch of your country home on a moonless autumn night. A breeze stirs the remaining leaves of the maple tree above you. Disturbed by your footfalls, some tiny creature rustles in the leaves on the ground. In the distance you hear a car, and from the sounds it makes, you judge that it is not a new car. It turns in at your driveway. You can't really see it—you are blinded by the glare of the headlights—but you hear the crunch of the gravel as it approaches. The car pulls to a stop. You hear the door open. "Hi, John," a voice says. "Hi, Mark," you reply. "Long time, no see."

We think of the world about us as the world we see, but hearing can play an important part in sensing and interpreting what is out there. The blind get along remarkably well by relying strongly on auditory cues. Their world must seem different from ours. We see photos and TV every day, and we regard these as depicting objects and events in a world familiar to us though sight. Indeed, like a camera, the eye has a lens that forms an image on the retina. Our visual world is built by interpreting the two images on our two retinas, and on the way they change as we move our eyes and head.

Bregman (1990) has written at length about the problems of auditory scene analysis. The ear has no lens. The sounds of the world reach our eardrums through two quarter-inch holes in our head. The sound waves that reach these holes are modified by reflections from the torso and the pinnae (outer ears). These modifications change as we move our head. Inside the ear, the sound waves in the auditory canal cause the eardrum to vibrate. That vibration is carried through a network of small bones to cause vibrations in the fluid-filled cochlea. The cochlea sorts them out according to the range of frequency. Some aspects of the time evolution of the oscillations are processed and compared. An example of this would be the comparisons of relative intensities and times of arrival at the two ears, through which we sense the direction of a sound source.

We may hope someday to have a complete and reliable model of human hearing, perhaps attained through guessing at the right neural networks and finding the values of parameters through means at hand. Beyond that, we hope someday to implement, or at least outline, an artificial sense of hearing that will hear as we do, with the same sensitivities and insensitivities that our own hearing exhibits. Then we will be on the road to a true virtual reality.

We do know that research has already disclosed a number of particular examples of things that we do and do not hear. Many of these are recounted in this chapter. It is good for composers and performers to remember them and keep them in mind. It is no use to produce fine gradations of sound that the auditory system cannot hear because of masking or because of the insensitivity of some component of the ear. It can be embarrassing to seek loudness in vain by piling up spectral components that mask each other. The ability of our ears to disregard or group together wanted sounds in the presence of unwanted sounds that could confuse us is essential to our perception of the world through sound.

8.2 The "Cocktail Party Effect"

Figure 8.1 has to do with distinguishing what one speaker is saying when another speaker is talking at the same time. This is easier when the two speakers stand in different locations than when they are in the same place. The ability to follow what one speaker says in the presence of the chatter of many others is called the "cocktail

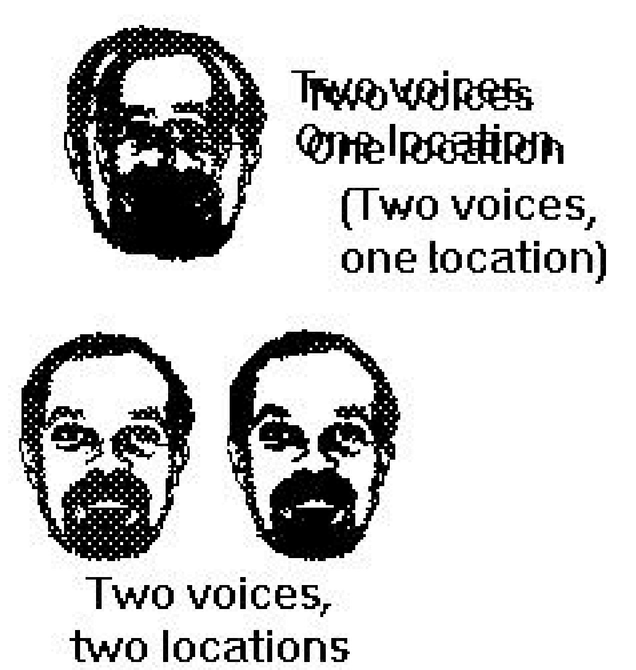

Figure 8.1 Two ears are better than one. We can listen better to one of two voices if the speakers are in different locations.

party effect." This is most easily illustrated to an audience by recording the voice of a single speaker reading different passages of text in succession.

When the different passages of text are simultaneously played at the same level over the same loudspeaker, we may be able to pick up a few words or phrases of one text or the other. It is impossible to follow either when both come from the same location.

It is quite different if we hear one voice from a loudspeaker to the left and the other from a loudspeaker to the right. We can to some degree "tune in" on either speaker, concentrating on the voice, left or right, that we wish to follow.

This "tuning in" is a consequence of the ear's ability to adjust the times of travel in neural pathways from the two ears so that the signals we wish to follow add up, and the interference doesn't. Whether or not we understand all the details of the process, we can see its remarkable consequences in simplified experiments.

8.3 Binaural Masking

Suppose that we use earphones to put various combinations of an identical noise and a sinusoidal signal into the left and right ears, as shown in figure 8.2

In figure 8.2 (a), at the top, the noise and the signal are the same in each ear. The subject cannot hear the signal.

In figure 8.2(b) the noise is the same in each ear, but the sign of the sinusoidal signal in one ear has been inverted (180 degree phase shift). The hearer can now hear the signal despite the noise. Somehow, the sense of hearing manages to make the sinusoidal signals, which arrive at the ears in different phases, add in phase, while the noise doesn't add in phase.

In 8.2(c) noise and signal go to one ear only. The noise keeps the subject from hearing the sinusoidal signal.

In 8.2(d) we have the same noise and signal in one ear as in (c) and the same noise, without signal, in the other ear. The subject can now hear the signal, for the addition of the noise in the other ear has enabled him to "filter" it out. The ear somehow shifts the combined response to the noise in the two ears so as to partially cancel the noise.

Such reduction of masking through binaural effects can be as large as 15 dB at 500 Hz; it falls to 2 or 3 dB for frequencies above 1500 Hz.

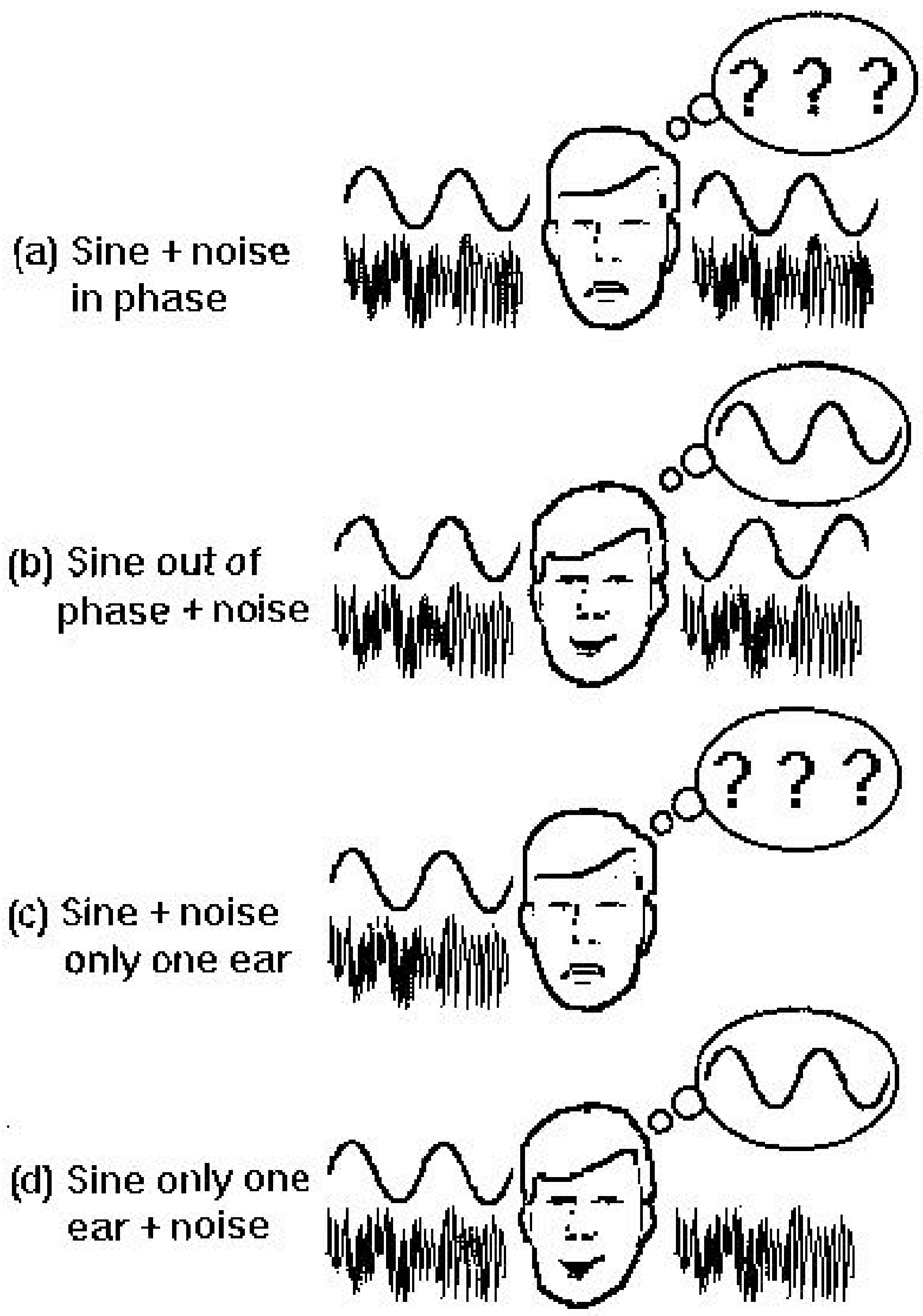

Figure 8.2 Binaural hearing of a sine wave obscured by noise. If the signal is different in the two ears and the noise is not, it is easier to hear the signal through the noise. With a signal to one ear only, noise in both ears is better than noise to the signal ear only. (Modified from Moore, 1964.)

This ability of the auditory system to shift the relative times at which the signals of the two ears are added is essential to the cocktail party effect.

8.4 The Precedence Effect

In figure 8.2 the "signal" is a steady sine wave, which is unrealistic. A sine wave goes on forever. Noises, clicks, voices, and musical tones have distinct beginnings, and these are essential to the effectiveness of our ability to hear sound from a source in a room that has reflecting walls. In such a room, the combination of many reflected sounds may reach us with greater intensity than direct sound, but that doesn't fool our ears. We hear a sound as coming from the direction from which it first reaches us. This is called the *precedence effect.*

The precedence effect can be demonstrated quite easily, and with startling clarity, as shown in figure 8.3. To do this, the same voice

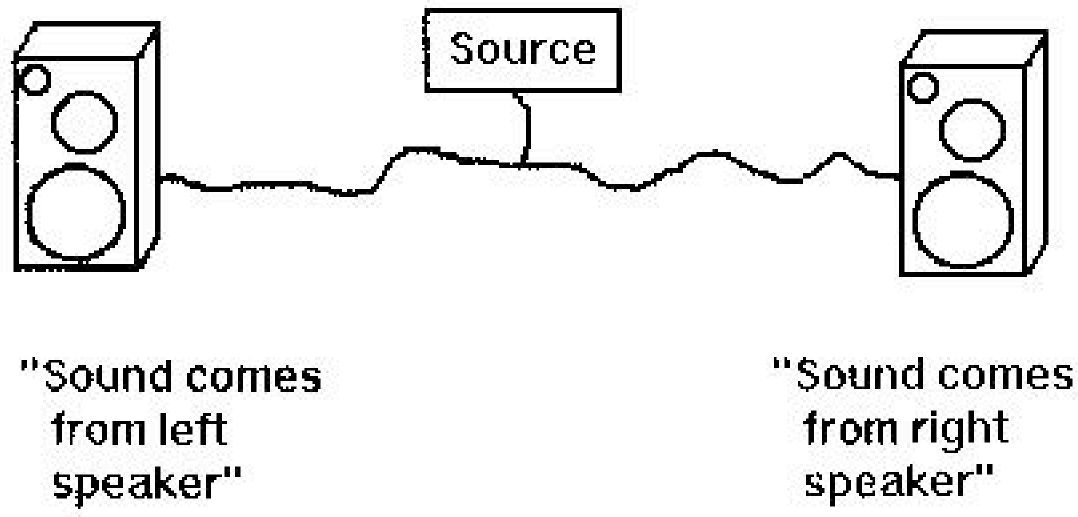

Figure 8.3 When the same signal comes from two loudspeakers, a listener hears it as coming from the nearer speaker. This is not a volume effect but is related to the arrival times of the sounds from the two speakers.

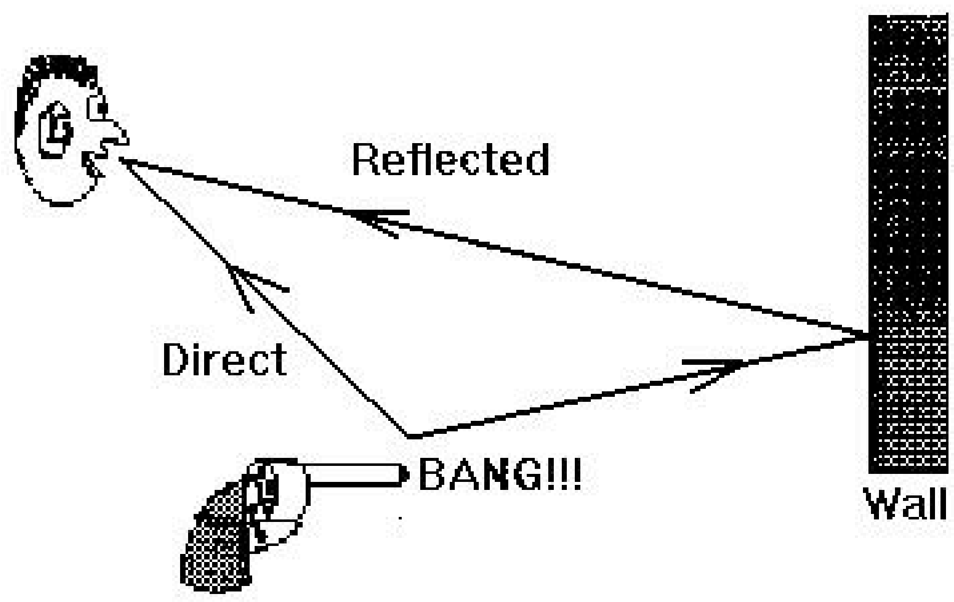

Figure 8.4 A sudden sound arriving both directly and reflected from a smooth wall is heard as having an echo only if the path of the reflected sound is 30 to 40 feet longer than the direct path.

signal is played at the same level from two loudspeakers. The voice asks members of the audience to raise their hands if they hear the voice as coming from the right-hand speaker, and then to raise their hands if they hear the voice as coming from the left-hand speaker.

Typically, those on the left-hand side of the room will hear the voice as coming from the left-hand speaker, and those on the right-hand side of the room will hear it as coming from the right-hand speaker, which is closer to them.

The precedence effect must be one of general sound arrival time rather than of exact waveform, for the behavior described above is unaltered if we reverse the phase of the signal to one of the speakers. Reversing the phase ruins the effectiveness of the stereo system at low frequencies, but the precedence effect persists.

The precedence effect has a long and interesting history that is recounted very well in a paper by Mark Gardner (1968b). One early experiment, described by Joseph Henry in 1849, is illustrated in figure 8.4. A listener hears a sharp sound, such as a hand clap or gunshot, both directly and reflected from a large, smooth wall. When

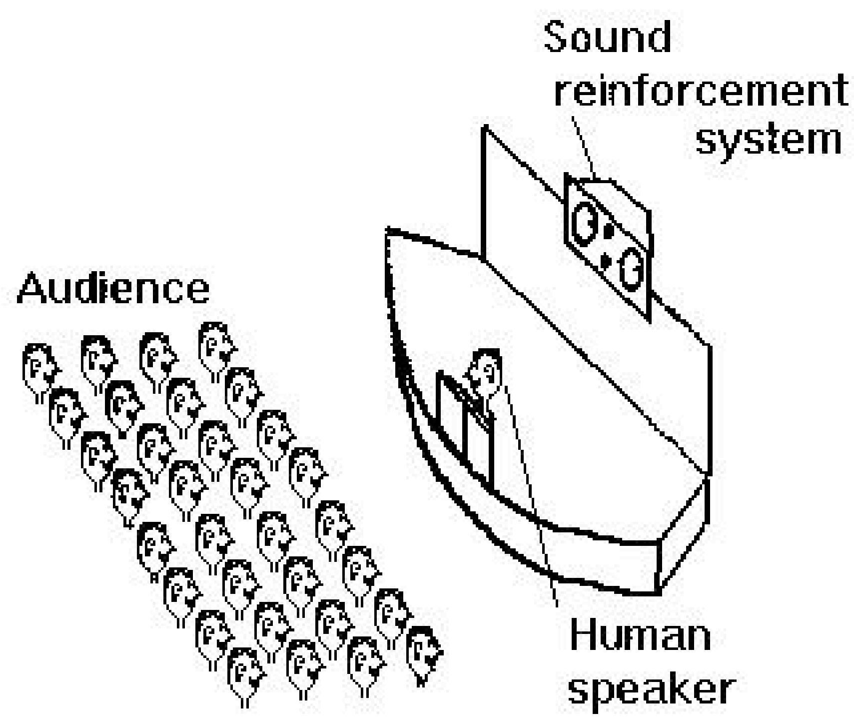

Figure 8.5 If a speaker's amplified voice comes from a speaker more distant from the audience than the speaker's mouth, the audience tends to hear the voice as coming from the speaker's mouth.

the source of the sound is moved away from the wall, at some distance the listener hears two sounds, the original hand clap and its echo. An echo was reported when the difference of distance of the direct sound vs. the source sound reflected from the wall was 30 to 40 feet, or a time difference of 1/40 to 1/30 of a second (25 to 35 milliseconds).

Figure 8.5 illustrates the use of the precedence effect in a public address system. The loudspeaker that emits the amplified voice of the speaker is farther from the audience than the speaker's mouth. Hence, if the amplification isn't too great, the audience tends to hear the speaker's voice as coming from the speaker's mouth rather than from the loudspeaker above, even though the sound coming from the loudspeaker may actually be louder than the speaker's voice. This is also reinforced by the visual cues of the speaker's moving lips and hand gestures.

The precedence effect makes the desired function of two-channel or stereo systems difficult. Ideally, we want a stereo system in which we hear the same source positions as we walk around in the room. But suppose the stereo channels are simply sounds picked up by two separated microphones or by a stereo microphone. The stereo channels will somehow share a component of a complicated sound from a single source. Through the precedence effect we tend to hear the complicated sound as coming from the loudspeaker nearer to us. Thus, we may hear a singing or speaking voice as coming from either the left or the right loudspeaker, even if the voice is located directly in front of us. However, disparities in sound intensity can be used to shift the sound toward the louder source.

Harris (1960) studied this effect with headphones for sinusoidal tones and for low-pass and high-pass clicks. The trade-off in decibels per millisecond time difference varies somewhat with the nature of the signal, so the use of a "left" or "right" centering control in a stereo system somewhat blurs the sense of direction, which is best for a central setting and a listener equidistant from the speakers.

The stereo channels for pop music are commonly obtained from single-channel recordings of individual instruments that are combined ingeniously to give an effect of sound localization or of being immersed in sound. One very simple way to assure a sort of immersion in sound is to filter out successive regions of the frequency spectrum and send alternate sections of the spectrum to the two stereo channels.

Various sorts of experiments give various critical time intervals that we can compare with the 50 or 60 milliseconds characteristic of the precedence effect, and the 30 or more milliseconds necessary for the perception of echoes. Rasch (1979) found that if players are not more than 30 to 50 milliseconds apart, they are perceived as playing "together." Hirsch (1959) and Patterson and Green (1970) found that when two half-second tones of different frequencies are played in succession, the order (low–high or high–low) can be heard if the time separation is more than about 20 milliseconds. But Patterson and Green also found that one can distinguish the difference in sound of a waveform played forward or backward for waveforms as short as 2 milliseconds. And, in estimating the direction of a sound source, Brian Moore (1989) showed that we can sense differences in direction that correspond to changes in relative time of arrival of the sound at our ears of as little as 10 or 20 microseconds.

The time range of some 50 to 60 milliseconds over which the precedence effect ties reflected sounds together without echo is comparable with our ability to hear differences in the time succession of musical notes rather than to our ability to make use of time differences in judging the direction of a sound source, or our ability to hear differences in the structure of short waveforms. It appears that the auditory system uses time disparity in different ways for different purposes.

8.5 Reverberation

Strong, discrete reflections of sound give echoes, which are a most objectionable phenomenon in bad concert halls. A multitude of suc-

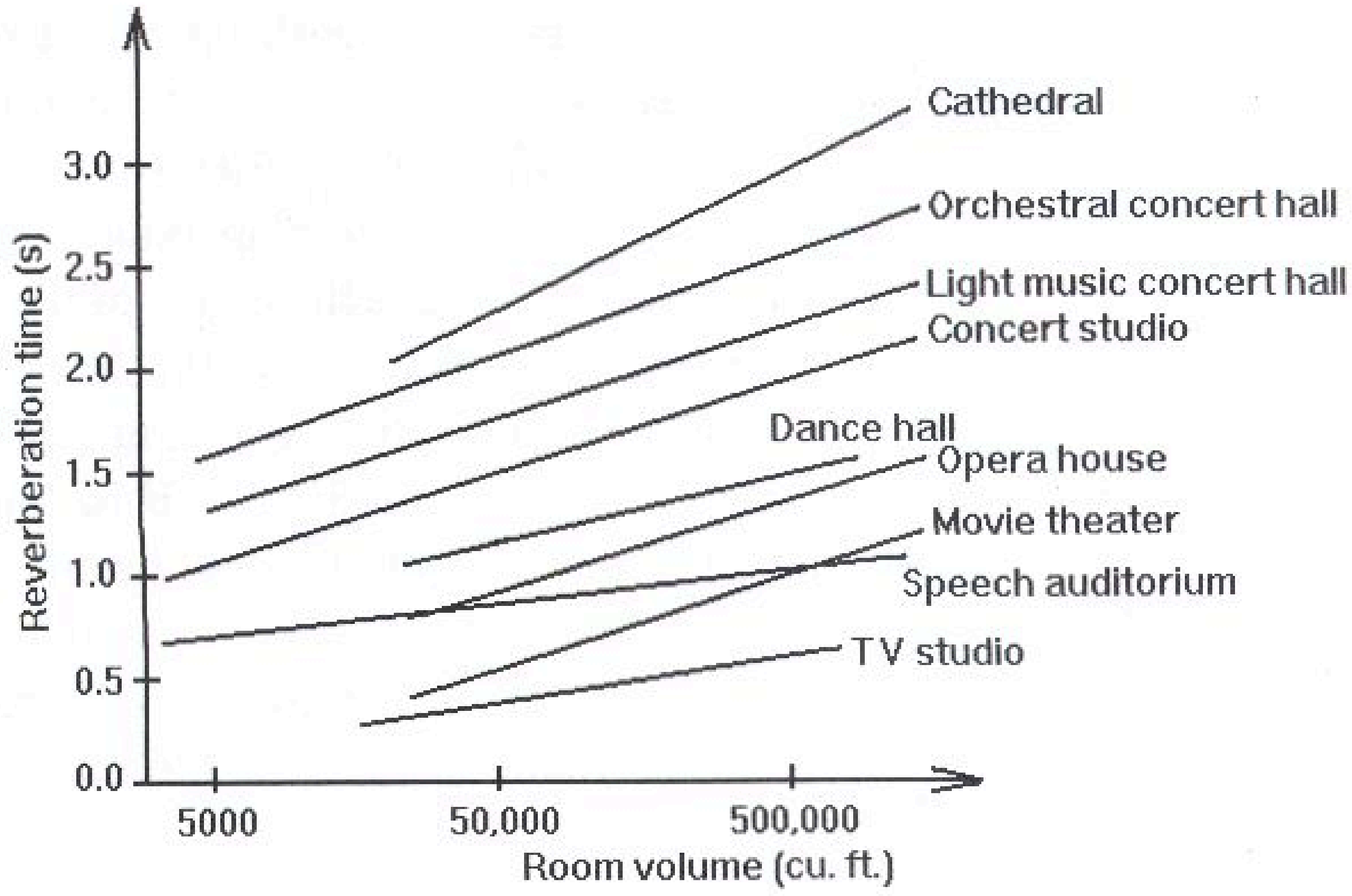

Figure 8.6 Appropriate reverberation times for enclosures of different sizes and purposes.

cessive reflections between walls that are very nearly parallel gives *flutter,* a phenomenon that "puts an edge on" voice and other complex sounds. But a multitude of small reflections that arrive at irregular times and die out smoothly constitute *reverberation,* without which music sounds dull, dry, and weak.

The reverberation in a room or hall is commonly specified by the number of seconds after which the sound produced by a short excitation has decayed by 60 dB. As indicated in figure 8.6, the optimum amount of reverberation for various musical purposes depends on the size of the enclosure and on the nature of the music. Organ music echos majestically through the stony spaces of a medieval cathedral, and so can singing if one doesn't need or want to attend to the words. More reverberation is acceptable in a concert hall than in an opera house, in which we wish to pay some attention to the details of the star's performance.

8.6 A Sense of Distance

Reverberation gives us a sense of some aspects of a room, such as the size and the materials that line the walls. It can also give us a clue to our distance from a sound source. In his book *Experiments in Hearing* (1980), Békésy notes that "In radio broadcasting it is well known that a listener in front of a loudspeaker can judge with considerable accuracy the distance of the speaker from the microphone. The greater the distance, the more reverberant are the sounds, and

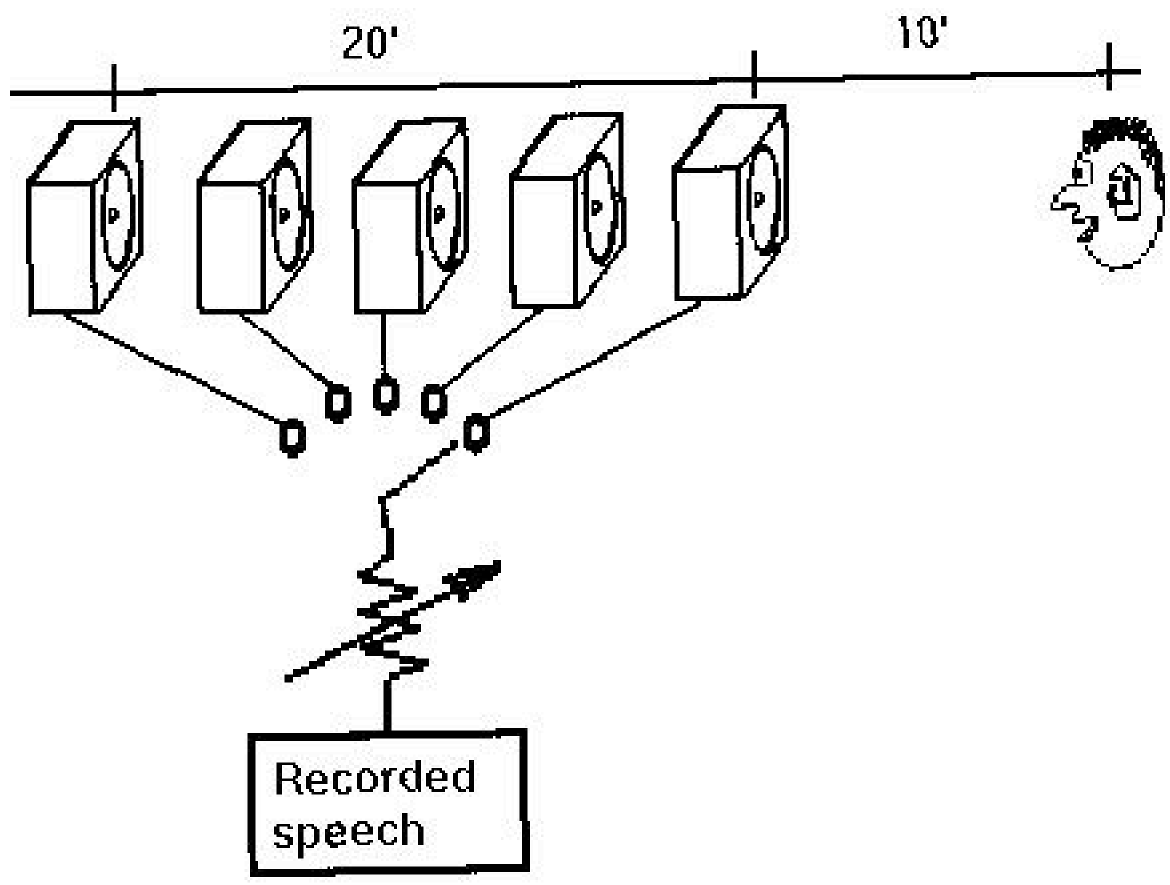

Figure 8.7 Speakers in an anechoic chamber. If intensity at the listener is the same, no matter which speaker the sound comes from, the listener will always judge it as coming from the nearest speaker.

the farther behind the loudspeaker they seem to be." Békésy references four papers published from 1927 to 1935.

Gardner (1968a) experimented with a voice from one of a sequence of visible loudspeakers in an anechoic chamber. As shown in figure 8.7, the speakers lay in the same direction from the listener, but at different distances. The levels at the speakers were adjusted so that the sound intensity at the listener's position was the same, no matter which speaker the voice came from. In the absence of reverberation, the listener always chose the nearest possible source as that of the origin of the voice. From our real-world experience, lack of reverberation means that the source is near the ear.

The addition of reverberation is a common way of controlling the apparent distance of computer-generated sound sources. Chowning (1971) and F. R. Moore (1983) have shown how reverberation and attenuation of high frequencies can be used to give clues to the apparent distance of a sound source. Usually, reverberation at the same level is used for all sound sources. Chowning also showed how the Doppler effect, the shift in frequency of a sound source moving toward or away from the hearer, can be used to give a sense of motion of a sound source.

8.7 Sense of Direction, Left or Right

Reverberation can give us a sense of the distance of a sound source, as well as a grand sense of being immersed in the sound of an orchestra. A sense of direction of a particular source is important in interpreting the sounds that reach our ears. It is through the precedence effect that we are able to sense the direction of a sound source in a closed space with reflecting walls.

In judging the direction of a sound source we make use of onsets or changes in sound that are characteristic of speech, musical notes, or noise. In a reverberant room it is nearly impossible to sense the direction of the source of a persistent sine wave.

Meaningful experiments concerning the effects of frequency on sensing the direction of a single sound source have been made by using clicks with various spectra, such as high-pass clicks or low-pass clicks. Our ability to sense the direction, left or right, of a sound source depends on the relative times of arrival and on relative intensities of the sound reaching the two ears.

The effects of differences in intensity in the two ear canals are greatest for high-frequency sounds, that is, sounds of short wavelength. Such sounds are reflected by the torso and shadowed by the head. The relative sound levels in the two ears depend, to some extent, on the direction that the head is turned. Chiefly the sound in the ear canal aimed toward the source is stronger than that in the ear canal facing away from the source.

The pinnae or outer ears cause a somewhat greater response to high-frequency sounds that are straight ahead than to high-frequency sounds from behind the head.

The relative times of arrival at the two ears are somewhat frequency dependent, but whatever the frequency, the sound arrives earlier at the ear nearer the source.

Experiments by Harris (1960), using headphones, show that there is an inverse relationship between the relative intensities at the two ears and the relative times of arrival at the two ears. This trade-off in decibels per millisecond is somewhat dependent on frequency.

For sounds with broad spectra, binaural localization in a stereo system is best when the two loudspeakers are the same distance from the listener. Through adjustment of the relative intensities of the signals in the two stereo channels, a listener who is not equidistant from the stereo speakers can hear a sound as coming from a point between the speakers, but the sense of direction may be fuzzy.

The accuracy of judgments of left or right can be amazing. Differences of around a degree can be detected. This corresponds to a disparity of only 10 microseconds in the time of arrival at the two ears.

But to get a sense that one musical note is played before or after another, the time disparity must be over a thousand times as great.

The sense of direction can be rendered fuzzy by putting the end of a finger in one ear. This has a greater effect on relative amplitude than on relative time of arrival. Fairly accurate left-or-right judgments can still be made with a finger in one ear.

8.8 Sense of Direction, Up or Down

Suppose someone makes a sound like a click or jingles keys. The location of the sound source is straight ahead of you but variable in height. In this case, there can be no difference in relative intensities at the two ears or in times of arrival. Yet a listener can tell whether the jingling or click comes from high or low. This is true for other sounds with significant high-frequency components.

Batteau (1967) pointed out that if one folds over his/her external ears, he/she completely loses the ability to tell high from low, that is, to localize high-pitched sounds in the median plane. Gardner (1973a) has published an excellent article on localization in the median plane. The chief effects of modifying the pinna occur for frequencies of 4000 Hz and higher.

Batteau thought of the effect of the pinna in terms of reflections of sound, of modifying its time waveform. The same effect can be thought of in terms of the contribution of the folds of the pinna to frequency response. Gardner (1973b) showed the variations with frequency of sound level in the auditory canal. Using a mannequin and a simulated ear, he investigated both a "normal" pinna and one whose convolutions had been occluded with a plasticine plug. He found that the crucial frequency responses appear to lie chiefly above 4,000 Hz.

There is a different effect of apparent height of a sound source in the median plane. Sounds with a strongly high-frequency spectrum can seem to have a source higher than sounds of a lower frequency spectrum.

My two stereo speakers used to be on the floor against the wall of the living room. Standing listeners heard recorded clapping sounds in Paul Lansky's composition "The Sound of Two Hands" (1992) as if the source were four or more feet from the floor, and occasional sounds of lower spectra seemed to come from the floor. This difference in apparent source height persisted when the pinnae were folded over.

8.9 Dummy Heads and Spatial Effects

Whatever spatial effects a two-channel, or even a four-channel, stereo system may provide, it cannot duplicate the entire sound field produced by one or several players in an enclosed space. Can we use headphones to produce in the two auditory canals of a listener the exact sound pressures that would reach the ears of a listener while sitting in a concert hall?

Schroeder, Gottlob, and Siebrasse (1974) addressed this problem by constructing a dummy head with appropriate pinnae and auditory canals. Microphones in the auditory canals provided two channels of sound.

By feeding the signal from the left microphone of the dummy head to a left headphone on a human listener, and feeding the signal from the right dummy microphone to the right listener headphone, we can approximate the signals in the auditory canal of a listener seated in the space in which the recording was made. Does the listener actually "hear" what he/she would have heard in the room?

The effect is better than listening to two ordinary stereo channels with headphones, but it is still not easy to get an adequate sense of a truly external sound source. If you were wearing headphones while listening to binaural signals from a dummy head, you might hear effects such as a nearby sound source behind you, or someone whispering into your ear, or the sound coming from a point within your head. Recording from one's own ear canals is better than recording from a dummy head. Some manufacturers of binaural dummy heads for research and music recording make casts of a specific set of ears, and these manufacturers will produce a binaural head with a custom set of pinnae.

Schroeder, Gottlob, and Siebrasse got more realistic results by using two loudspeakers to produce, at a given head position, the sound intensities that the left and right ears of the dummy head would have "heard" at a given head position. When this is done, the apparent source position can be to the left of the left speaker or to the right of the right speaker. This technique was used in comparing the same recorded performance as it would be heard when played on the stages of various European concert halls.

8.10 Our Perception of Noise

A great deal that is important in the hearing of music has been learned through experiments with sine waves, even though pure sine waves don't qualify as musical sounds. Gaussian noise really isn't a musical sound either, but we can learn important things through studying the perception of noise.

"Gaussian noise" is a sequence of random events smoothed by a special type of low-pass filtering. Suppose that the probability that an extremely short pulse of unit amplitude will occur in a very short time interval dt is p^*dt. There will be an average number p of pulses per second, but the times of their occurrence will be unpredictable. If we low-pass-filter this sequence of pulses with a filter of bandwidth B, and if the number of pulses per second is large compared with $2B$, the filtered waveform will approach a Gaussian noise.

The bandwidth of a compact disc is around 20,000 Hz. We might think that for a sequence of pulses to be heard as noise, the number of pulses per second would have to be large compared with 40,000 samples per second. What do we actually hear as the pulse rate is increased?

Up to a few tens of pulses per second, we hear separate, unpredictable pulses, as in a Geiger counter. At a few hundred pulses per second we hear an erratic signal, clearly "pulsey," but we can't detect all pulses individually. Above a few thousand (say, 5000) pulses per second, we hear a "smooth" noise. Our hearing runs individual pulses together so that we no longer hear any pulse as a separate event.

This "running together" of pulses is associated with the critical bandwidth of the ear, not with the overall frequency range of hearing. Though differences in time of as little as 10 microseconds are detectable in sensing left and right in binaural hearing, random sequences of pulses run together as smooth noise when the average pulse spacing is around 0.2 millisecond, or 200 microseconds.

8.11 In Conclusion

Unlike the eye, the ear affords no direct image of the world around us. Yet, in addition to our ability to judge frequency and distance, our binaural hearing can give us an amazing amount of help in dealing with that world. Our two ears enable us to listen in a direction, partially disregarding sounds coming from other directions. We can judge vertical as well as horizontal direction of a sound source, even when the initial sound is followed by echoes from the walls of a room. Distinct echoes tell us about reflecting surfaces. A host of small reflections adds a warm quality to the sound, and also tells us

Table 8.1 Discrimination of the ear

Accuracy (just noticeable differences)
Intensity, about 1 dB
Frequency, about 10 cents (0.6 percent)
Bandwidth
Around 20,000 Hz
Critical bandwidth, around f/5 (a musical minor third)
Time resolution
Binaural azimuth, around 20 microseconds (not heard as time order, before or after)
Clicks, around 2 milliseconds (sounds different with different time order, but no sense of before or after)
Musical tones, around 20 milliseconds (sense of before or after)
Playing together (simultaneity), around 30–50 milliseconds
Echoes (clapping, voice), 50–60 milliseconds (precedence effect)
Threshold of hearing
At 3,000 Hz, around 10 watts per square meter for people with acute hearing
Dynamic range
90–120 dB at 3,000 Hz
30–60 dB at 30 Hz
Masking
Around 30 dB, but a complex set of phenomena

something of the distance of the source. Understanding such aspects of sound and our hearing can help computer musicians and composers, both in generating appropriate sounds and in evaluating and using performance spaces.

In exercising the art of music it is good to have a rough sense of some simple limitations of hearing. I have brought together a number of these limitations in table 8.1.

References

Batteau, D. W. (1967). "The Role of the Pinna in Human Localization." *Proceedings of the Royal Society of London*, B168 (1011): 159–180.

Begault, D. (1994). *3D Audio for Virtual Reality and Multimedia Applications* San Diego: Academic Press.

Békésy, G. von (1980). *Experiments in Hearing*. Huntington, N.Y.: Robert E. Krieger. (1st ed. New York: McGraw-Hill, 1960).

Bregman, A. S. (1990). *Auditory Scene Analysis*. Cambridge, Mass.: MIT Press.

Chowning, J. M. (1971). "The Simulation of Moving Sound Sources." *Journal of the Audio Engineering Society*, 199, 2–6.

Gardner, M. B. (1968a). "Proximity Effect in Sound Localization." *Journal of the Acoustical Society of America*, 43, 163–169.

———. (1968b). "Historical Background of the Haas and/or Precedence Effect." *Journal of the Acoustical Society of America*, 43, 1243–1248.

———. (1973a). "Some Monaural and Binaural Factors in Median Plane Localization." *Journal of the Acoustical Society of America,* 54. 1489–1495.

———. (1973b). "Some Single- and Multiple-Source Localization Effects." *Audio Engineering,* 21, 430–437.

Gordon, J. W. (1983). "The Perceptual Synchrony of Attack Times of Musical Tones." *Journal of the Acoustical Society of America,* 82. 88–105.

Harris, G. G. (1960). "Binaural Interactions of Impulsive and Pure Tones." *Journal of the Acoustical Society of America,* 32, 685–692.

Hirsch, I. C. (1959). "Auditory Perception of Temporal Order." *Journal of the Acoustical Society of America,* 31, 759–767.

Lansky, P. (1992). "The Sound of Two Hands." On the compact disc, *Homebrew:* Bridge Records.

Moore, B. C. J. (1989). *An Introduction to the Psychology of Hearing.* San Diego: Academic Press.

Moore, F. R. (1983). "A General Model for Spatial Processing of Sounds." *Computer Music Journal* (Fall) 7, 6–15.

Patterson, J. H., and D. M. Green. (1970). "Discrimination of Transient Signals Having Identical Energy Spectra." *Journal of the Acoustical Society of America,* 48, 121–131.

Rasch, R. A. (1979). "Synchronization in Performed Ensemble Music." *Acustica,* 43, 121–131.

Resnick, S. B., and Feth, L. L. (1975). "Discriminability of Time-Reversed Pairs of Clicks." *Journal of the Acoustical Society of America,* 57, 1493–1499.

Ronkin, D. A. (1970). "Monaural Detection of Phase Difference Between Clicks." *Journal of the Acoustical Society of America,* 47, 1091–1099.

Schroeder, M. R., D. Gottlob, and K. F. Siebrasse (1974). "Comparative Study of European Concert Halls. Correlation of Subjective Preference with Geometric and Acoustic Parameters." *Journal of the Acoustical Society of America,* 56, 1195–1201.