# 1

# Sound and the auditory system

This chapter provides a brief introduction to the physical nature of sound, the manner in which it is transmitted and transformed within the ear, and the nature of auditory neural responses.

## The nature of auditory stimuli

The sounds responsible for hearing consist of rapid changes in air pressure that can be produced in a variety of ways – for example, by vibrations of objects such as the tines of a tuning fork or the wings of an insect, by puffs of air released by a siren or our vocal cords, and by the noisy turbulence of air escaping from a small opening. Sound travels through the air at sea level at a velocity of about 335 meters per second, or 1,100 feet per second, for all but very great amplitudes (extent of pressure changes) and for all waveforms (patterns of pressure changes over time). Special interest is attached to periodic sounds, or sounds having a fixed waveform repeated at a fixed frequency. Frequency is measured in hertz (Hz), or numbers of repetitions of a waveform per second; thus, 1,000 Hz corresponds to 1,000 repetitions of a particular waveform per second. The time required for one complete statement of an iterated waveform is its period. Periodic sounds from about 20 through 16,000 Hz can produce a sensation of pitch and are called tones. For reasons to be discussed shortly, it is generally considered that the simplest type of periodic sound is a sine wave or pure tone (shown in Figure 1.1A), which has a sinusoidal change in pressure over time. A limitless number of other periodic waveforms exists, including square waves (Figure 1.1B) and pulse trains (Figure 1.1C). Periodic sounds need not have simple, symmetrical waveforms: Figure 1.1D shows a periodic sound

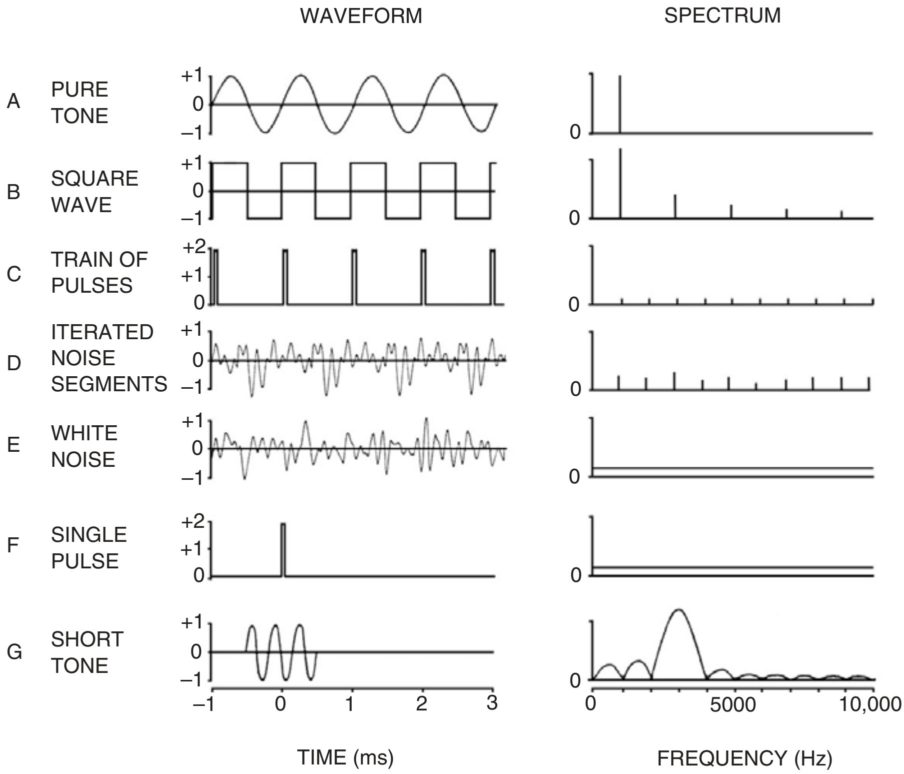


**Figure 1.1** Waveforms and amplitude spectra. The waveforms A through E continue in time to produce the spectra as shown. Periodic waveforms A through D have line spectra, the others either continuous spectra (E and F), or a band spectrum (G). See the text for further discussion.

produced by iteration of a randomly generated waveform. The figure also depicts the waveforms of some nonperiodic sounds: white or Gaussian noise (Figure 1.1E), a single pulse (Figure 1.1F), and a short tone or tone burst (Figure 1.1G).

The waveforms shown in Figure 1.1 are time-domain representations in which both amplitude and time are depicted. Using a procedure developed by Joseph Fourier in the first half of the nineteenth century, one can represent any periodic sound in terms of a frequency-domain or spectral analysis in which a sound is described in terms of a harmonic sequence of sinusoidal components having appropriate frequency, amplitude, and phase relations. (Phase describes the portion of the period through which a waveform has advanced relative to an arbitrarily fixed reference time.) A sinusoidal or pure tone consists of a single spectral component, as shown in Figure 1.1A. The figure also shows the power spectra corresponding to the particular complex (nonsinusoidal) periodic sounds shown in Figures 1.1B, 1.1C, and 1.1D. Each of these sounds has a period of one millisecond, a fundamental frequency of 1,000 Hz (corresponding to the waveform repetition frequency), and harmonic components corresponding to integral multiples of the 1,000 Hz fundamental as indicated.

Frequency analysis is not restricted to periodic sounds: nonperiodic sounds also have a spectral composition as defined through use of a Fourier integral or Fourier transform (for details see Hartmann, 1998). Nonperiodic sounds have either continuous or band spectra rather than line spectra, as shown for the sounds depicted in Figures 1.1E, 1.1F, and 1.1G.

As we shall see, frequency analysis of both periodic and nonperiodic sounds is of particular importance in hearing, chiefly because a spectral analysis is performed within the ear leading to a selective stimulation of the auditory nerve fibers.

Although Figure 1.1 shows how particular waveforms can be analyzed in terms of spectral components, it is also possible to synthesize waveforms by adding together sinusoidal components of appropriate phase and amplitude. Figure 1.2 shows how a sawtooth waveform may be approximated closely by the mixing of only six harmonics having appropriate amplitude and phase.

The range of audible amplitude changes is very large. A sound producing discomfort may be as much as $10^6$ times the amplitude level at threshold. Sound level can be measured as power as well as by amplitude or pressure at a particular point. Power usually can be considered as proportional to the square of the amplitude, so that discomfort occurs at a power level $10^{12}$ times the power threshold. The term ''sound intensity'' is, strictly speaking, the sound power arriving from a specified direction, and passing through a unit area perpendicular to that direction. However, the term ''intensity'' is often used interchangeably with ''power,'' although the latter term has no directional specificity.

In order to span the large range of values needed to describe the levels of sound normally encountered, a logarithmic scale has been devised. The logarithm to the base 10 of the ratio of a particular sound power level to a reference power level defines the level of the sound in Bels (named in honor of Alexander Graham Bell). However, the Bel is a rather large unit, and it is conventional to use a unit 1/10 this size, the decibel (or dB) to express sound levels. The level in dB can be defined as:

$$\mathrm{dB} = 10 \log_{10} (I_1/I_2)$$

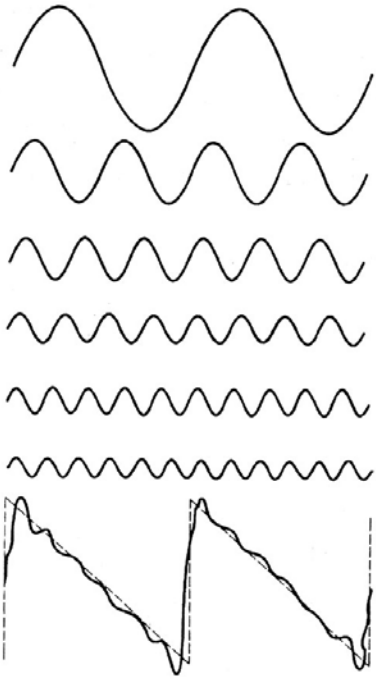


**Figure 1.2** Synthesis of a complex waveform through addition of harmonically related sinusoidal components. The approximation of a sawtooth waveform could be made closer by the addition of higher harmonics of appropriate amplitude and phase. (From Brown and Deffenbacher, 1979.)

where $I_1$ is the intensity or power level of the particular sound of interest, and $I_2$ is the reference level expressed as sound intensity. One can also calculate decibels on the basis of pressure or amplitude units using the equation:

$$\mathrm{dB} = 20 \log_{10} (P_1/P_2)$$

where $P_1$ is the relative pressure level being measured and $P_2$ is the reference pressure level. The standard reference pressure level is 0.0002 dyne/cm$^2$ (which is sometimes expressed in different units of 20 micropascals). The level in dB measured relative to this standard is called the Sound Pressure Level (or SPL). Sound-level meters are calibrated so that the numerical value of the SPL can be read out directly. There is another measure of sound level, also expressed in dB, called the Sensation Level (SL), which is used occasionally in psychoacoustics. When measuring SL, the level corresponding to the threshold of a sound for an individual listener is used as the reference level rather than the standard physical value employed for SPL, so that dB SL represents the level above an individual's threshold. Since SL is used relatively infrequently, dB will always refer to SPL unless otherwise specified.

To give some feeling for sound pressure levels in dB, the threshold of normal listeners for sinusoidal tones with frequencies between 1,000 and 4,000 Hz (the range exhibiting the lowest thresholds) is about 0 dB (the standard reference level); the ambient level (background noise) in radio and TV studios is about 30 dB, conversational speech about 65 dB, and the level inside a bus about 90 dB. Some rock bands achieve levels of 120 dB, which approaches the threshold for pain and can cause permanent damage to hearing following relatively brief exposures.

Experimenters can vary the relative intensities of spectral components by use of acoustic filters which, in analogy with light filters, pass only desired frequency components of a sound. A high-pass filter transmits only frequency components above a lower limit, and a low-pass filter transmits only frequencies below an upper limit. Bandpass filters (which transmit frequencies within a specified range) and band-reject filters (which block frequencies within a specified range) are available. The characteristics of high-pass and low-pass filters can be expressed in terms of both cut-off frequency (conventionally considered as the frequency at which the filter attenuation reduces power by half, or 3 dB), and the slope, or roll-off, which is usually expressed as dB/octave beyond the cut-off frequency (a decrease of one octave corresponds to halving the frequency). Bandpass filters are characterized by their bandwidth (the range in hertz between the upper and lower cut-off frequencies), and they can also be characterized by their ''Q'' (the bandwidth divided by the center frequency of the filter). In neurophysiological work, $Q_{10}$ is sometimes used in which 10 dB downpoints are used to express the bandwidth rather than the conventional value of 3 dB. Filter types are shown in Figure 1.3.

## Our auditory apparatus

*The outer ear and the middle ear*

It is convenient to consider the ear as consisting of three divisions. The outer ear, also called the pinna (plural ''pinnae'') or auricle, is shown in Figure 1.4. It appears to contribute to localization of sound sources by virtue of its direction-specific effect on the intensity of certain frequency components of sounds, as will be discussed in the next chapter. The human pinna is surrounded by a simple flange (the helix) which is extended considerably in some other mammals to form a conical structure functioning as a short version of the old-fashioned ear trumpet. These acoustic funnels can enhance sensitivity to high frequency sounds when pointed toward their source by controlling muscles, as well as being of help in locating the sound source.

After the acoustic transformation produced by reflections within our pinna, the sound passes through the ear canal (or external auditory meatus) which

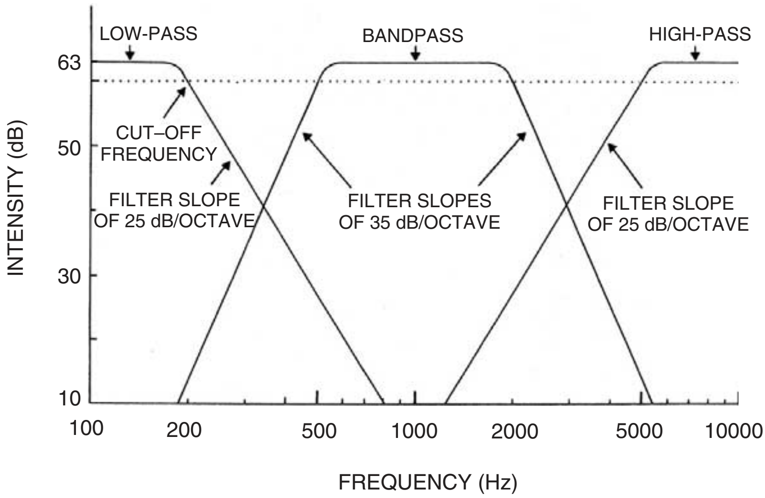


**Figure 1.3** Characteristics of filters. Low-pass, high-pass, and bandpass filters are illustrated, along with filter slopes (dB/octave) and cut-off frequencies (frequencies at which there is a 3 dB reduction in intensity).

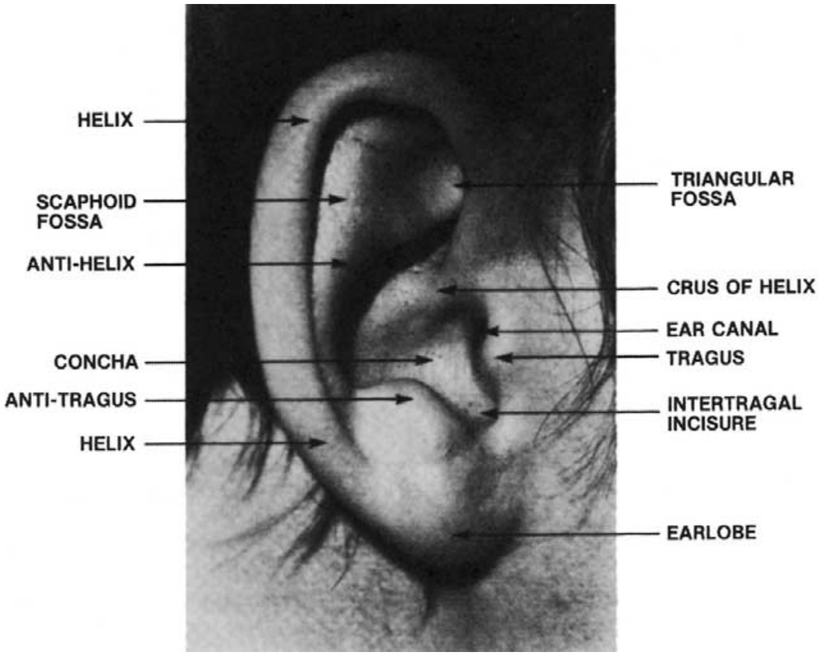


**Figure 1.4** The outer ear (other names: pinna and auricle). The major anatomical features are shown.

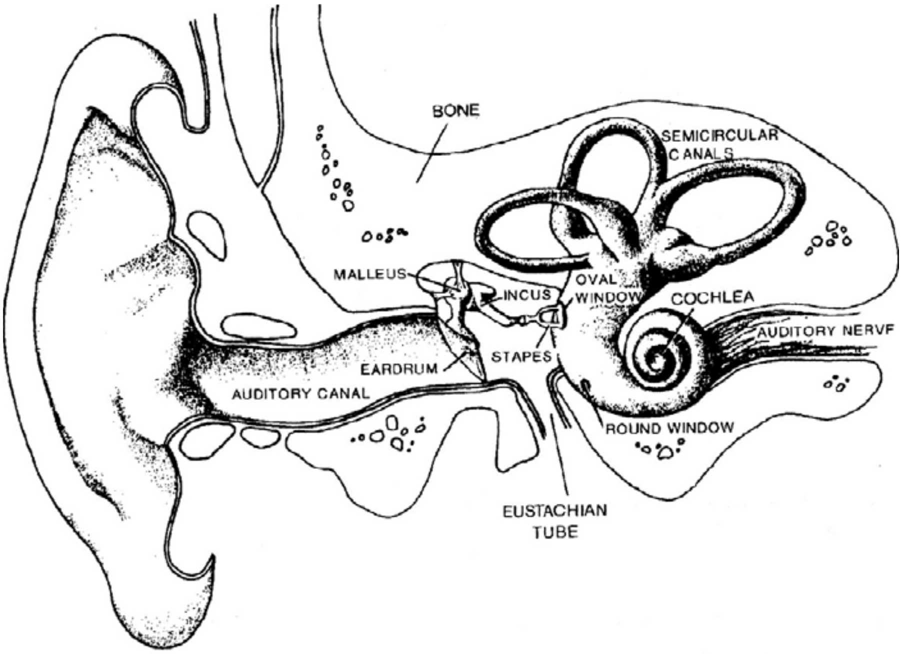


**Figure 1.5** Diagram of the entire ear. The outer, middle, and inner ears are shown, along with adjacent structures. (Adapted from Lindsay and Norman, 1977.)

ends at the eardrum, or tympanum, as shown in Figure 1.5. This canal is more than a passive conduit. Its length is roughly 2.5 cm, and it behaves in some respects like a resonant tube, such as an organ pipe. The effect of this resonance is to amplify frequencies appreciably (5 dB or more) from about 2,000 through 5,500 Hz, with a maximum amplification of about 11 dB occurring at about 4,000 Hz (Wiener, 1947). The pressure changes at the end of the canal cause the eardrum (or tympanum) to vibrate. This vibration is picked up and transmitted by a chain of three small bones (or ossicles) located in the middle ear. The first of these bones, the malleus (or hammer) is attached to the tympanum, and its movement is transmitted to the incus (or anvil) and thence to the stapes (or stirrup). The stapes is connected to the oval window at the base of the fluid-filled cochlea. This window forms a boundary separating the middle and inner ear. The passage of sound through the cochlea is shown in Figure 1.6, and will be discussed subsequently.

The middle ear permits the airborne sound to be converted to liquid borne sound without the great loss which would otherwise occur. When sound in air impinges directly upon a liquid, a loss of about 30 dB (99.9 percent of the power) takes place, with most of the sound being reflected back into the air. Three physical principles act to increase the efficiency of the transmission of

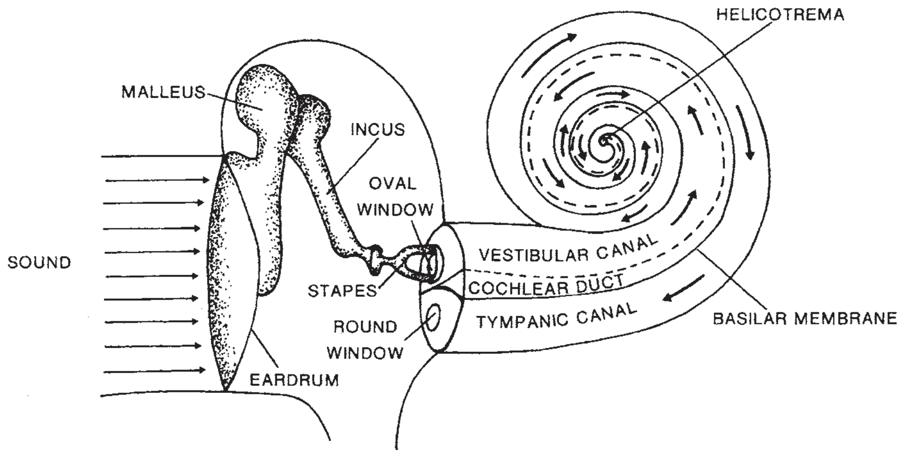


**Figure 1.6** Conversion from air-conduction to liquid-conduction of sound by the ear. For details see the text. (Adapted from Lindsay and Norman, 1977.)

sound by the middle ear: (1) the curvature of the tympanum (which is somewhat conical in shape) causes it to act as a more efficient mechanical transformer (Tonndorf and Khanna, 1972); (2) the chain of three ossicles acts like a lever with a small mechanical advantage; and (3) the force applied to the larger area of the tympanic membrane, when transmitted to the much smaller area of the footplate of the stapes embedded in the oval window, produces a considerable mechanical advantage. This last factor is the most important of the three.

There are two muscles within the middle ear that can lessen the intensity of very strong stimuli and minimize the possibility of damage to the inner ear. One of these, the tensor tympani muscle, is attached to the malleus, and the other, the stapedius muscle, is attached to the stapes. These muscles are sometimes compared in their effect to the iris of the eye – a high level of stimulus intensity causes a reflex contraction of the muscles resulting in a decrease in stimulation. Once the threshold for initiating the reflex is reached, there is an attenuation up to about 0.6 or 0.7 dB for each 1 dB increase in the stimulus, with a maximum attenuation of perhaps 30 dB for low frequency sounds (the reduction in intensity is greatest for frequencies below 1,000 Hz). Middle ear muscle contraction can also reduce distortions which would otherwise occur from overloading the ossicular chain. A very few people can contract their middle ear muscles voluntarily, and thus attenuate even relatively faint sounds at will. For most of us the action is strictly reflexive, either in response to an external sound of 80 dB or more, or as an action that precedes the self-generation of sound in speaking or chewing food. The reflex activity of these muscles in response to external sound is very quick, perhaps 10 ms for very intense sounds, but this still cannot protect against sudden harmful sounds such as gunshots.

Are the intra-aural muscles more than an analog of the eye's iris? There have been some interesting speculations. Lawrence (1965) suggested that since animal studies have indicated that muscle activity is to some degree independent in the two ears, intermittent monaural changes in intensity and phase produced by muscle contraction can help in directing attention to sources at different azimuths under noisy conditions. Simmons (1964) considered that low frequency sounds produced by eating might mask high frequency environmental sounds of importance and that selective attenuation of these self-generated sounds by the intra-aural muscles could release these external sounds from masking and permit their detection.

*Structure of the inner ear*

The inner ear contains not only the receptors responsible for hearing, but also receptors involved in detecting acceleration and maintaining balance. The tortuous anatomical structure of the inner ear has led to its being called the labyrinth. The vestibule of the labyrinth contains the utricle and saccule which appear to be sensitive to linear acceleration of the head and to orientation in the gravitational field. There are also three bony semicircular canals, each canal lying in a plane that is at right angles to the other two, much as the floor and two adjacent walls of a room form three mutually perpendicular planes. This arrangement of the canals permits detection of the components of rotary acceleration in any of the planes (see Figure 1.5).

The bony spiral structure within the inner ear called the cochlea (from the Latin name for snail) contains the organ for hearing. This coiled tube consists of about 2.5 turns and has a length of about 3.5 cm. It is partitioned into three canals or ducts called scalae. Two of the scalae are joined: the scala vestibuli or vestibular canal (which has at its basal end the flexible oval window to which the stapes is attached) communicates (via a small opening called the helicotrema at the apex of the spiral) with the scala tympani or tympanic canal (which has the flexible round window at its basal end). These two scalae contain a fluid called perilymph, and when the oval window is flexed inward by the stapes, the almost incompressible perilymph causes the round window to flex outward. As shown in Figure 1.7, the scala vestibuli is bounded by Reissner's membrane and the scala tympani by the basilar membrane. Between these two membranes lies the scala media or cochlear duct, which has a closed end near the helicotrema and which contains a fluid called endolymph. A third fluid called cortilymph is found within the tunnel of Corti.

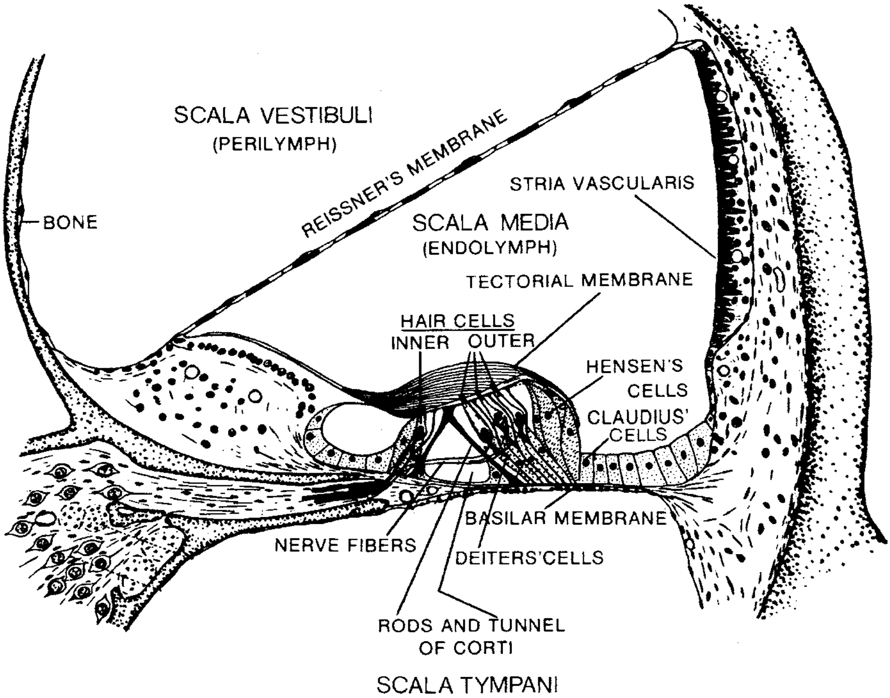


**Figure 1.7** Cross-section of the cochlea, showing the organ of Corti and associated structures. This diagram is based on the guinea pig, but is representative of the human inner ear as well. (Adapted from Davis, Benson, Covell, *et al.*, 1953.)

These fluids have different ionic compositions and different electric potentials which appear to play a role in the processes eventuating in the stimulation of the auditory nerve. A complex neuroepithelium called the organ of Corti lies on the basilar membrane and contains the hair cells that are responsible for stimulating the auditory nerve. As shown in Figure 1.8, the hair cells consist of two types: the outer hair cells, which are closer to the cochlear wall, and the inner hair cells, each topped by a cuticular plate containing stereocilia. The stereocilia are bathed in endolymph, whereas most of the receptor cell is surrounded by the cortilymph found in the extracellular spaces of the organ of Corti. There are about 12,000 outer hair cells, each containing about 100 to 200 stereocilia arranged in parallel rows. Within these rows, resembling a letter V or W, the longest stereocilia have their tips attached to the tectorial membrane (the consequences of this coupling of the basilar and tectorial membranes will be discussed shortly). There are about 3,500 inner hair cells forming a single row in the shape of a flattened letter U. Each inner hair cell contains about 50 stereocilia having tips that lie without attachment in a groove called Hensen's stripe on the underside of the tectorial membrane. The stereocilia of both the