1 Physiological Aspects of Cochlear Hearing Loss

I INTRODUCTION

Hearing loss caused by damage to the cochlea is probably the most common form of hearing loss in the developed countries. Its most obvious symptom, and the one that is almost always assessed in the clinic, is an elevation of the threshold for detecting sounds. However, it is also accompanied by a variety of other changes in the way that sound is perceived. Even if sounds are amplified (e.g. by a hearing aid) so that they are well above the threshold for detection, the perception of those sounds is usually abnormal; the person with cochlear hearing loss often reports that the sounds are unclear and distorted, and that it is hard to hear comfortably over a wide range of sound levels. A common complaint is difficulty in understanding speech, especially when background sounds or reverberation are present. One of the main aims of this book is to explain why these problems occur and why current hearing aids are of limited benefit in compensating for the problems.

The book assumes that the reader has a basic knowledge of physics and acoustics, for example an understanding of what is meant by terms such as sinusoid, spectrum, frequency component and the decibel. The reader who is not familiar with these terms should consult a textbook, such as An Introduction to the Psychology of Hearing (Moore, 2003) or Signals and Systems for Speech and Hearing (Rosen and Howell, 1991). For the reader who knows these things, but needs a reminder, many of the key terms are defined briefly in the Glossary. Most of the terms that appear in italics in the text are defined in the Glossary. One concept that may not be familiar is that of a linear system. This topic is of importance, since the normal peripheral auditory system shows significant nonlinearities, whereas the system becomes more linear when cochlear damage occurs. Hence, this chapter starts with a description of the properties of linear and nonlinear systems. It then goes on to consider the physiology and the function of the normal and damaged cochlea.

II LINEAR AND NONLINEAR SYSTEMS

The auditory system is often thought of as a series of stages, the output of a given stage forming the input to the next. Each stage can be considered as a device or system, with an input and an output. For a system to be linear, certain relationships between the input and output must hold true. The following two conditions must be satisfied:
1. If the input to the system is changed in magnitude by a factor $k$, then the output should also change in magnitude by a factor $k$, but be otherwise unaltered. This condition is called **homogeneity**. For example, if the input is doubled, then the output is doubled, but without any change in the form of the output. Thus, a plot of the output as a function of the input would be a straight line passing through the origin (zero input gives zero output) – hence the term **linear system**. Such a plot is called an **input-output function**. An example of such a function is given in panel (a) of Figure 1.1.

2. The output of the system in response to a number of independent inputs presented simultaneously should be equal to the sum of the outputs that would have been obtained if each input were presented alone. For example, if the response

---

**Figure 1.1.** The left column shows input-output functions for a linear system (a), a nonlinear system with ‘hard’ peak clipping (d) and a nonlinear system with more progressive ‘saturation’ (g). The middle column shows outputs from these systems in response to a sinusoidal input. The third column shows spectra of the outputs.
to input A is X and the response to input B is Y, then the response to A and B together is simply X + Y. This condition is known as superposition.

When describing a system as linear, it is usually assumed that the system is time-invariant. This means that the input-output function does not change over time. For example, if the input is \( I \) and the output is \( O \), the relationship between the input and the output would be:

\[
O = cI,
\]

where \( c \) is a constant that does not vary with time.

When a sinusoid is used as an input to a linear system, the output is a sinusoid of the same frequency. This is illustrated in panel (b) of Figure 1.1. The amplitude and phase of the output may, however, be different from those of the input. Assume that the input is a single sinusoid whose waveform as a function of time, \( I(t) \), can be described by:

\[
I(t) = A \sin(2\pi ft),
\]

where \( A \) is the peak amplitude of the input and \( f \) is the frequency in Hz (cycles per second). The output as a function of time, \( O(t) \), could then be represented by:

\[
O(t) = G \times A \sin(2\pi ft + \phi),
\]

where \( G \) is a constant representing the amplitude ratio between the input and output, and \( \phi \) is a constant representing the phase shift between the input and output. \( G \) is sometimes referred to as the gain of the system.

Since the output of a linear system in response to a sinusoidal input is itself sinusoidal, the spectrum of the output, by definition, consists of a single frequency component. This is illustrated in panel (c) of Figure 1.1. More generally, the output of a linear system never contains frequency components that were not present in the input signal. The response of a linear system may, however, vary with the frequency of the input sinusoid. This is equivalent to saying that the constants \( G \) and \( \phi \) in Equation 1.3 can vary with the input frequency.

In practice, many devices or systems are linear as long as the input is not too large. Excessive inputs may cause the system to become nonlinear; more details are given later on. Such a system is usually called linear, even though it can become nonlinear under extreme conditions. As an example, consider a loudspeaker. The input is a voltage and the output is a movement of the cone of the loudspeaker, which can produce audible sound waves. For the types of inputs that are typically used for a loudspeaker, the response is approximately linear; the conditions of homogeneity and superposition are obeyed. If the input voltage varies in a sinusoidal manner, the movement of the cone is almost sinusoidal. However, if the frequency of the input is changed, holding the magnitude of the input constant, the magnitude of the movement of the cone may
vary. In this case, we would say that the loudspeaker does not have a ‘flat’ frequency response; \( G \) may vary with frequency. Similarly, \( \phi \) may vary with frequency. Other examples of systems that usually operate in a nearly linear way are microphones, amplifiers and the output transducers used in hearing aids (often called receivers).

When waveforms other than sinusoids are applied as the input to a linear system, the output waveform will often differ from that of the input. For example, if the input to a linear system is a square wave, the output is not necessarily a square wave. This is one reason for the popularity of sinusoids in auditory research; sinusoids are the only waveforms which are always ‘preserved’ by a linear system. If a system is linear, then it is relatively easy to predict its output for any arbitrary complex input. As a first step, the output is measured as a function of frequency for a sinusoidal input. Essentially, the values of \( G \) and \( \phi \) are determined as a function of the input frequency. To predict the output for a complex input, a Fourier analysis of the input is performed. This gives a description of the input in terms of the amplitudes and phases of its sinusoidal components. The output for each of the sinusoidal components comprising the input can then be calculated. Finally, using the principle of superposition, the output in response to the whole complex can be calculated as the sum of the outputs in response to its individual sinusoidal components. This is a powerful method, and it gives another reason for using sinusoids as stimuli.

As mentioned earlier, many linear systems become nonlinear if the input is made large enough. An example is shown in panel (d) of Figure 1.1. The input-output function is linear over a large part of its range, but it flattens out for large positive or negative values of the input. This is sometimes called saturation or peak clipping and can occur in systems such as transistor amplifiers and condenser microphones. In the example shown, the clipping is symmetrical, in that it occurs at the same absolute value for positive and negative values of the input. When a sinusoid is used as input to such a system and the peak amplitude of the sinusoid, \( A \), is sufficiently large, the output is no longer sinusoidal. This is illustrated in panel (e) of Figure 1.1. The output is periodic, with the same period as the input sinusoid, but the waveform is distorted. In this case, the output contains frequency components that are not present in the input. This is illustrated in panel (f) of Figure 1.1. In general, the output of a nonlinear system in response to a single sinusoid at the input contains one or more components (sinusoids) with frequencies that are integer multiples of the frequency of the input. These components are referred to as harmonics, and the nonlinear system is said to introduce harmonic distortion. For example, if the input was a sinusoid with a frequency of 500 Hz, the output might still contain a component with this frequency, but components with other frequencies might be present too, for example 1000, 1500, 2000... Hz.

Another example of a nonlinear input-output function is shown in panel (g) of Figure 1.1. In this case, the function does not show ‘hard’ clipping, but the slope becomes more shallow when the absolute value of the input or output exceeds a certain value. This type of input-output function can occur in valve (tube) amplifiers, moving coil microphones and loudspeakers, and it can also occur in the auditory system. The output waveform, shown in panel (h) of Figure 1.1, is less distorted than
when hard clipping occurs, and the output spectrum, shown in panel (i), reveals less harmonic distortion.

If the input to a nonlinear system consists of two sinusoids, then the output may contain components with frequencies corresponding to the sum and difference of the two input frequencies, and their harmonics, as well as the original sinusoidal components that were present in the input. These extra components are said to result from intermodulation distortion. For example, if the input contains two sinusoids with frequencies $f_1$ and $f_2$, the output may contain components with frequencies $f_1 - f_2, f_1 + f_2, 2f_1 - f_2, 2f_2 - f_1$ and so on. These components are referred to as intermodulation distortion products, and, in the case of the auditory system, they are also called combination tones.

When a system is nonlinear, the response to complex inputs cannot generally be predicted from the responses to the sinusoidal components comprising the inputs. Thus, the characteristics of the system must be investigated using both sinusoidal and complex inputs.

Often, the input and output magnitudes of a system are plotted on logarithmic axes (the decibel scale is an example). In that case, the input and output magnitudes are not specified as instantaneous values (e.g. as the instantaneous voltage in the case of an electrical signal). Generally, instantaneous magnitudes can have both positive and negative values, but it is not possible to take the logarithm of a negative number. Instead, the input and output are averaged over a certain time, and the magnitude is expressed as a quantity which cannot have a negative value. Typically, both the input and output are specified in terms of their power (related to the mean-square value) or their root-mean-square value. Sometimes, the peak amplitude may be used. For a linear system, the condition of homogeneity still applies to such measures. For example, if the input power is doubled, the output power is also doubled. When plotted on ‘log-log’ axes, the input-output function of a linear system is a straight line with a slope of unity. To see why this is the case, we take the logarithm of both sides of Equation 1.1 ($O = cI$). This gives:

$$\log(O) = \log(cI) = \log(c) + \log(I)$$  \hspace{1cm} (1.4)$$

The value of $\log(c)$ is itself a constant. Therefore, since $\log(O)$ is simply equal to $\log(I)$ plus a constant, the slope of the line relating $\log(O)$ to $\log(I)$ must be unity. In a nonlinear system, the slope of the input-output function on logarithmic axes differs from unity. Say, for example, that the output is proportional to the square of the input:

$$O = cI^2$$  \hspace{1cm} (1.5)$$

Taking logarithms of both sides gives:

$$\log(O) = \log(cI^2) = \log(c) + 2\log(I)$$  \hspace{1cm} (1.6)$$
In this case, the slope of the input-output function on logarithmic axes is two. When the slope of the input-output function is greater than one, the nonlinearity is referred to as expansive. If the output were proportional to the square-root of the input \((O = cI^{0.5})\), the slope of the function would be 0.5. When the slope of the input-output function is less than one, the nonlinearity is referred to as compressive. Examples of input-output functions plotted on log-log axes will be presented later in this chapter, in connection with the response of the basilar membrane (BM) within the cochlea.

III STRUCTURE AND FUNCTION OF THE OUTER AND MIDDLE EAR

Figure 1.2 shows the structure of the peripheral part of the human auditory system. It is composed of three parts, the outer, middle and inner ear. The outer ear includes the pinna and the auditory canal, or meatus. The pinna and meatus together create a broad resonance which enhances sound levels at the eardrum, relative to those obtained in the absence of the listener’s head, over the frequency range from about 1.5 to 5 kHz.

Figure 1.2. Illustration of the structure of the peripheral auditory system showing the outer, middle and inner ear. Redrawn from Lindsay and Norman (1972).
This is illustrated in Figure 1.3. The maximum boost is typically about 12–15 dB in the region around 2.5 kHz. When a hearing aid is fitted so as to block the meatus, this broad resonance is lost for a behind-the-ear aid, reduced in magnitude for a concha aid, and shifted to higher frequencies for a completely-in-the-canal aid. This has to be taken into account when fitting hearing aids and especially when adjusting the gain as a function of frequency.

At medium and high frequencies, the sound reaching the eardrum is significantly modified by the pinna, head and upper torso. Specifically, when the sound contains a broad range of frequencies, the pinna introduces a complex pattern of peaks and notches into the spectrum. This pattern varies systematically with the direction of the sound source relative to the head, and the spectral patterns thus provide important information about the location of sound sources (see Chapter 7 for more details).

Sound travels down the meatus and causes the eardrum, or tympanic membrane, to vibrate. The eardrum forms the outer boundary of the middle ear. These vibrations are transmitted through the middle ear by three small bones, the ossicles, to a membrane-covered opening in the bony wall of the spiral-shaped structure of the inner ear – the cochlea. This opening is called the oval window and it forms the outer boundary of the middle ear. The three bones are called the malleus, incus and stapes (popularly
known as the hammer, anvil and stirrup), the stapes being the lightest and smallest of these and the one which actually makes contact with the oval window.

The major function of the middle ear is to ensure the efficient transfer of sound energy from the air to the fluids in the cochlea. If the sound were to impinge directly onto the oval window, most of it would simply be reflected back, rather than entering the cochlea. This happens because the resistance of the oval window to movement is very different from that of air. This is described as a difference in acoustical impedance. The middle ear acts as an impedance-matching device or transformer that improves sound transmission and reduces the amount of reflected sound. This is accomplished mainly by the 27:1 ratio of effective areas of the eardrum and the oval window, and to a small extent by the lever action of the ossicles. Transmission of sound energy through the middle ear is most efficient at middle frequencies (500–5000 Hz), which are the ones most important for speech perception; see Chapter 8. This is illustrated in Figure 1.4, adapted from Glasberg and Moore (2006), which shows an estimate of the relative effectiveness of transmission through the middle ear as a function of frequency.

The ossicles have minute muscles attached to them which contract when we are exposed to intense sounds. This contraction, known as the *middle ear reflex* or *acoustic reflex*, is probably mediated by neural centres in the brain stem (Liberman and Guinan, 1998). The reflex can be triggered by sound of any frequency, but it reduces the transmission of sound through the middle ear only at low frequencies (below about 1000 Hz). It may help to prevent damage to the delicate structures inside the cochlea. However, the activation of the reflex is too slow to provide any protection against impulsive sounds, such as gunshots or hammer blows.

![Figure 1.4](image)

**Figure 1.4.** The transfer function of the middle ear, plotted as relative response versus frequency. The response was arbitrarily labelled as 2.6 dB at 1 kHz. The estimate comes from Glasberg and Moore (2006).
For moderate sound levels (below about 90 dB SPL), the outer and middle ear behave essentially as linear systems; they do not introduce significant harmonic or intermodulation distortion. However, at high sound levels both the tympanic membrane and the ossicles may vibrate in a nonlinear manner, and the acoustic reflex also introduces nonlinearity. These nonlinearities may result in audible harmonic and intermodulation distortion.

IV STRUCTURE AND FUNCTION OF THE NORMAL COCHLEA

IV.1 THE COCHLEA, THE BASILAR MEMBRANE AND THE ORGAN OF CORTI

The inner ear is also known as the cochlea. It is shaped like the spiral shell of a snail. However, the spiral shape does not appear to have any functional significance, and the cochlea is often described as if the spiral had been ‘unwound’. The cochlea is filled with almost incompressible fluids, and it has bony rigid walls. It is divided along its length by two membranes, Reissner’s membrane and the BM (see Figure 1.5).

Figure 1.5. Cross-section of the cochlea, showing the BM, Reissner’s membrane and the organ of Corti. Redrawn from Davis (1962).
The start of the spiral, where the oval window is situated, is known as the base; the other end, the inner tip, is known as the apex. It is also common to talk about the basal end and the apical end. At the apex there is a small opening (the helicotrema) between the BM and the walls of the cochlea which connects the two outer chambers of the cochlea, the scala vestibuli and the scala tympani. Inward movement of the oval window results in a corresponding outward movement in a membrane covering a second opening in the cochlea – the round window. Such movements result in pressure differences between one side of the BM and the other (i.e. the pressure is applied in a direction perpendicular to the BM), and this results in movement of the BM (see below for details). The helicotrema eliminates any pressure differences between the scala vestibuli and the scala tympani at very low frequencies. This prevents the BM from moving significantly in response to movements of the oval window caused by jaw movements or by slow changes in air pressure (such as occur when changing altitude). The helicotrema also reduces movement of the BM in response to low-frequency sounds.

On the side of the cochlea closest to the outer wall (the right-hand side in Figure 1.5), there is a structure called the stria vascularis. This plays a strong role in the metabolism of the cochlea and in creating the voltages (electrical potentials) that are essential for the normal operation of the cochlea. The stria vascularis is sometimes colloquially described as the ‘battery’ of the cochlea.

A third membrane, called the tectorial membrane, lies above the BM, and also runs along the length of the cochlea. Between the BM and the tectorial membrane are hair cells, which form part of a structure called the organ of Corti (see Figures 1.5 and 1.6). They are called hair cells because they appear to have tufts of

![Figure 1.6. Cross section of the organ of Corti as it appears in the basal turn of the cochlea.](image-url)
hairs, called stereocilia, at their apexes. The hair cells are divided into two groups by an arch known as the tunnel of Corti. Those on the side of the arch closest to the outside of the spiral shape are known as outer hair cells (OHCs), and they are arranged in three rows in cats and up to five rows in humans, although the rows are often somewhat irregular in humans (Wright et al., 1987). The hair cells on the other side of the arch form a single row, and are known as inner hair cells (IHCs). The stereocilia on each OHC form a V- or W-shaped pattern, and they are arranged in rows (usually about three) that are graded in height, the tallest stereocilia lying on the outside of the V or W. The stereocilia on each IHC are also arranged in rows graded in height, but the arrangement is more like a straight line or a broad arc. In humans, there are about 12 000 OHCs (per ear), each with about 140 stereocilia protruding from it, while there are about 3500 IHCs, each with about 40 stereocilia.

The tectorial membrane, which has a gelatinous structure, lies above the hair cells. It appears that the stereocilia of the OHCs actually make contact with the tectorial membrane, but this may not be true for the IHCs. The tectorial membrane appears to be effectively hinged at one side (the left in Figure 1.6). When the BM moves up and down, a shearing motion is created; the tectorial membrane moves sideways (in the left–right direction in Figure 1.6) relative to the tops of the hair cells. As a result, the stereocilia at the tops of the hair cells are moved sideways. The movement occurs via direct contact in the case of the OHCs, but in the case of the IHCs it may be produced by the viscous drag of fluid streaming between the upper part of the organ of Corti and the tectorial membrane. The movement of the stereocilia of the IHCs leads to a flow of electrical current through the IHCs, which in turn leads to the generation of action potentials (nerve spikes) in the neurones of the auditory nerve. Thus, the IHCs act to transduce mechanical movements into neural activity.

The IHCs and OHCs have very different functions. The great majority of afferent neurones, which carry information from the cochlea to higher levels of the auditory system, connect to IHCs; each IHC is contacted by about 20 neurones (Spoendlin, 1970). Thus, most information about sounds is conveyed via the IHCs. The main role of the OHCs is actively to influence the mechanics of the cochlea. The OHCs have a motor function, changing their length, shape and stiffness in response to electrical stimulation (Ashmore, 1987; Yates, 1995), and they can therefore influence the response of the BM to sound. The OHCs are often described as being a key element in an active mechanism within the cochlea. The function of this active mechanism is described in more detail below.

The action of the OHCs is partly under the control of higher centres of the auditory system. There are about 1800 efferent nerve fibres that carry information from the auditory system to the cochlea, most of them originating in the superior olivary complex of the brain stem. Many of these efferent fibres make contact with the OHCs, and can affect their activity (Liberman and Guinan, 1998). Thus, even the earliest stages in the analysis of auditory signals are partly under the control of higher centres.
IV.2 TUNING ON THE BASILAR MEMBRANE

When the oval window is set in motion by a sound, a pressure difference occurs between the upper and lower surface of the BM. The pressure wave travels almost instantaneously through the incompressible fluids of the cochlea. Consequently, the pressure difference is applied essentially simultaneously along the whole length of the BM. This causes a pattern of motion to develop on the BM. The pattern does not depend on which end of the cochlea is stimulated. Sounds which reach the cochlea via the bones of the head rather than through the air do not produce atypical responses.

The response of the BM to stimulation with a sinusoid takes the form of a travelling wave which moves along the BM from the base towards the apex. The amplitude of the wave increases at first and then decreases rather abruptly. The basic form of the wave is illustrated in Figure 1.7, which shows schematically the instantaneous displacement of the BM for four successive instants in time, in response to a low-frequency sinusoid. The four successive peaks in the wave are labelled 1, 2, 3 and 4. This figure also shows the line joining the amplitude peaks, which is called the envelope. The envelope shows a peak at a particular position on the BM.

The response of the BM to sounds of different frequencies is strongly affected by its mechanical properties, which vary progressively from base to apex. At the base the BM is relatively narrow and stiff. This causes the base to respond best to high frequencies. At the apex the BM is wider and much less stiff, which causes the apex to respond best to low frequencies. Each point on the BM is tuned; it responds best (with greatest displacement) to a certain frequency, called the characteristic frequency (CF), or best frequency, and responds progressively less as the frequency

![Figure 1.7](image_url)

**Figure 1.7.** The solid lines show the instantaneous displacement of the BM at four successive instants in time (labelled 1–4), derived from a cochlear model. The pattern moves from left to right, building up gradually with distance, and decaying rapidly beyond the point of maximal displacement. The dashed line represents the envelope traced out by the amplitude peaks in the waveform.
is moved away from the CF. It is now believed that the tuning of the BM arises from two mechanisms. One is referred to as the *passive mechanism*. This depends on the mechanical properties of the BM and surrounding structures, and it operates in a roughly linear way. The other is the *active mechanism*. This depends on the operation of the OHCs, and it operates in a nonlinear way. The active mechanism depends on the cochlea being in good physiological condition, and it is easily damaged. The travelling wave shown in Figure 1.7 is typical of what is observed when only the passive mechanism is operating.

Figure 1.8 shows the envelopes of the patterns of vibration for several different low-frequency sinusoids (data from von Békésy, 1960). Sounds of different frequencies produce maximum displacement at different places along the BM, that is, there is a frequency-to-place transformation. If two or more sinusoids with different frequencies are presented simultaneously, each produces maximum displacement at its appropriate place on the BM. In effect, the cochlea behaves like a frequency analyser, although with less than perfect resolution. The resolution is often described in terms of the sharpness of tuning. This refers to the ‘narrowness’ of the response patterns on the BM. In the case of responses to a single tone, as shown in Figure 1.7, it refers to the spread of the response along the BM; sharp tuning would be associated with a narrow spread.

Most of the pioneering work on patterns of vibration along the BM was done by von Békésy (1960). The vibration patterns found by von Békésy were rather broad; for example, the pattern for a 400-Hz sinusoid extended along almost the whole length of the BM (see Figure 1.8). However, these patterns probably reflect only the passive mechanism. The active mechanism would not have been functioning in von Békésy’s experiments, for two reasons. Firstly, he had to use very high sound levels – about 140 dB SPL; such high levels are known to damage the active mechanism. Secondly, he used cadaver ears, and the active mechanism ceases to function after death.

Recent work measuring BM responses to sound differs from that of von Békésy in several ways. Firstly, living animals have been used. Great care has been taken to keep the animals in good physiological condition during the measurements and to minimize trauma caused by the necessary surgery. Secondly, the techniques themselves are designed to be minimally invasive. Finally, rather than measuring the response of several different points on the BM to a single frequency, measurements have usually been made of the responses of a single point to sinusoids of differing frequency. In this case, the sharpness of tuning is often measured by adjusting the level at each frequency to produce a fixed response on the BM. If a point on the BM is sharply tuned, then the sound level has to be increased rapidly as the frequency is moved away from the CF. If the tuning is broad, then the sound level has to be increased only gradually as the frequency is moved away from the CF. The results show that the sharpness of tuning of the BM depends critically on the physiological condition of the animal; the better the condition, the sharper is the tuning (Khanna and Leonard, 1982; Sellick, Patuzzi and Johnstone, 1982; Leonard and Khanna, 1984; Robles, Ruggero and Rich, 1986; Ruggero, 1992; Robles and Ruggero, 2001).

The health of the cochlea is often monitored by placing an electrode in or near the auditory nerve, and measuring the combined responses of the neurones to tone bursts
Figure 1.8. Envelopes of patterns of vibration on the BM for a number of low-frequency sounds. Solid lines indicate the results of actual measurements, while the dashed lines are von Békésy’s extrapolations. Redrawn from von Békésy (1960).

or clicks; this response is known as the *compound action potential* (AP), or CAP. The lowest sound level at which an AP can be detected is called the AP threshold. Usually, the BM is sharply tuned when the AP threshold is low, indicating that the cochlea is in good physiological condition and the active mechanism is functioning.
Figure 1.9. Tuning curves measured at a single point on the BM. Each curve shows the input sound level required to produce a constant velocity on the BM, plotted as a function of stimulus frequency. The curve marked by open circles was obtained at the start of the experiment when the animal was in good physiological condition. The curve marked by filled squares was obtained after the death of the animal. Data from Sellick, Patuzzi and Johnstone (1982).

An example is given in Figure 1.9, which shows the input sound level (in dB SPL) required to produce a constant velocity of motion at a particular point on the BM, as a function of stimulus frequency (Sellick, Patuzzi and Johnstone, 1982). This is sometimes called a constant velocity tuning curve. It is not yet clear whether the effective stimulus to the IHCs is BM vibration amplitude (equivalent to displacement) or BM velocity. However, for a given frequency, velocity is directly proportional to displacement: the greater the amplitude, the faster the movement. At the start of the experiment, when AP thresholds were low, a very sharp tuning curve was obtained (open circles). This curve reflects the contribution of both the passive and the active mechanisms. As the condition of the animal deteriorated, the active mechanism ceased to function. The tuning became broader, and the sound level required to produce the criterion response increased markedly around the tip. The broad tuning curve recorded after death (filled squares) reflects the tuning produced by the passive mechanism alone.

The frequency at the tip of the tuning curve (the CF – where the sound level was lowest) shifted downwards when the condition of the animal deteriorated. This occurred because, at least for high CFs, the active mechanism gives maximum gain for a
frequency that is somewhat above the best frequency determined by the passive mechanism. In a healthy ear, the CF is determined mainly by the active mechanism. When the active mechanism ceases to function, the sharply tuned tip of the tuning curve is lost and the CF shifts to a lower frequency determined by the passive mechanism.

Even in a healthy ear, the balance between the active and passive mechanisms can change with sound level. If the input level of a sinusoid is held constant, and its frequency is varied, then the response of the BM (velocity or amplitude) at a specific point shows a peak for a specific frequency. However, the frequency which gives the maximum response often varies with the input sound level (Ruggero et al., 1997). For high CFs, this frequency decreases with increasing sound level, as the relative contribution of the active mechanism decreases. This implies that, for a fixed input frequency, the place on the BM showing the maximum response shifts towards the base with increasing sound level. Usually, the CF is specified for a low input sound level.

In summary, in a normal healthy ear each point along the BM is sharply tuned, responding with high sensitivity to a limited range of frequencies and requiring higher and higher sound intensities to produce a response as the frequency is moved outside that range. The sharp tuning and high sensitivity reflect the active process mediated by the OHCs.

IV.3 THE NONLINEARITY OF INPUT-OUTPUT FUNCTIONS ON THE BASILAR MEMBRANE

In a normal healthy ear, the response of the BM is nonlinear; when the input magnitude is increased, the magnitude of the response does not grow directly in proportion to the magnitude of the input (Rhode, 1971; Rhode and Robles, 1974; Sellick, Patuzzi and Johnstone, 1982; Robles, Ruggero and Rich, 1986; Ruggero, 1992; Ruggero et al., 1997; Robles and Ruggero, 2001). This is illustrated in Figure 1.10, which shows input-output functions of the BM for a place with a CF of 8 kHz (from Robles, Ruggero and Rich, 1986). A series of curves is shown; each curve represents a particular stimulating frequency, indicated by a number (in kHz) close to the curve. The output (velocity of vibration) is plotted on a logarithmic scale as a function of the input sound level (in dB SPL – also a logarithmic scale). If the responses were linear, the functions would be parallel to the dashed line. Two functions are shown for a CF tone (8 kHz), one (at higher levels) obtained about one hour after the other. The slight shift between the two was probably caused by a deterioration in the condition of the animal. An ‘idealized’ function for a CF tone, with the ordinate scaled in dB units (i.e. as 20\log_{10}(velocity)) is shown in Figure 1.11.

While the function for the CF tone is almost linear for very low input sound levels (below 20–30 dB) and approaches linearity at high input sound levels (above 90 dB), the function has a very shallow slope at mid-range levels. This indicates a compressive nonlinearity: a large range of input sound levels is compressed into a smaller range of responses on the BM. The form of this function can be explained in the following way. At low and medium sound levels the active mechanism amplifies the response on the BM. The amplification may be 50 dB or more (Robles and Ruggero, 2001).
At very low sound levels, below 20–30 dB, the amplification is roughly constant and is at its maximal value. As the sound level increases, the amplification progressively reduces. Thus, the response grows more slowly than it would in a linear system. When the sound level is sufficiently high, around 90 dB SPL, the active mechanism is unable to contribute any amplification, and the response becomes linear (although some researchers have reported a shallow growth of response even at very high sound levels). Hence, at high levels, the ‘passive’ response becomes dominant.

The nonlinearity mainly occurs when the stimulating frequency is close to the CF of the point on the BM whose response is being measured. For stimuli with

**Figure 1.10.** Input-output functions for a place on the BM with CF = 8 kHz. The stimulating frequency, in kHz, is indicated by a number close to each curve. The dashed line indicates the slope that would be obtained if the responses were linear (velocity directly proportional to sound pressure). Redrawn from Robles, Ruggero and Rich (1986).

**Figure 1.11.** Schematic input-output function of the BM for a sinusoid at CF. A decibel scale is used for both axes. The ordinate is scaled (arbitrarily) so that an input of 100 dB gives an output of 100 dB. The dashed line shows the slope that would be obtained if the response were linear.
frequencies well away from the CF, the responses are more linear. Hence, the curves for frequencies of 7 and 9 kHz (close to CF) show shallow slopes, while the curves for frequencies below 7 kHz and above 9 kHz show steeper (linear) slopes. Effectively, the compression occurs only around the peak of the response pattern on the BM. As a result, the peak in the distribution of vibration along the BM flattens out at high sound levels, which partly accounts for the broad tuning observed by von Békésy (1960).

Given that the BM response at moderate sound levels is highly nonlinear, one might expect that, in response to a single sinusoid, harmonics would be generated and the waveform on the BM would be significantly distorted. In fact, this does not seem to happen (Cooper and Rhode, 1992; Ruggero et al., 1997). The reason why is not fully understood. Perhaps the active mechanism involves feedback onto the BM in such a way that potential harmonic distortion is filtered out by the passive mechanism. However, in response to inputs containing more than one sinusoidal component, significant distortion can occur. This is described later on in this chapter.

IV.4 TWO-TONE SUPPRESSION

Experiments using two tones have revealed another aspect of nonlinearity on the BM, namely two-tone suppression. The effect is analogous to an effect that was first discovered from measurements of the responses of single neurones in the auditory nerve (see below for details). The response to a tone (called the ‘probe’ tone) with frequency close to the CF of the place on the BM being studied can be reduced by a second tone with a higher or lower frequency, especially when the second tone is higher in level than the probe tone (Rhode, 1977; Patuzzi, Sellick and Johnstone, 1984; Ruggero, Robles and Rich, 1992). The effect is illustrated in Figure 1.12 for a probe tone at 8.6 kHz (close to the CF) and a suppressor tone at 10.6 kHz. Ruggero, Robles and Rich (1992) provided a detailed comparison of the properties of mechanical two-tone suppression on the BM and two-tone suppression measured in the auditory nerve. They concluded that all of the properties match qualitatively (and mostly quantitatively) and that two-tone suppression in the auditory nerve probably originates from two-tone suppression on the BM.

IV.5 COMBINATION TONE GENERATION

Another aspect of BM nonlinearity is the generation of distortion products in response to two or more sinusoidal inputs. These products are often called combination tones. When two sinusoids are presented simultaneously, and their frequency separation is not too great, their response patterns overlap on the BM. It appears that, at the point of overlap, distortion products are generated that behave like additional sinusoidal tones. For example, if the two primary tones presented to the ear have frequencies $f_1$ and $f_2$ ($f_2 > f_1$), the distortion products have frequencies such as $2f_1 - f_2$ and $f_2 - f_1$. The distortion products appear to propagate along the BM to the locations tuned to their own frequencies (Robles, Ruggero and Rich, 1991). For example, the combination
Figure 1.12. An example of two-tone suppression for a place on the BM with CF = 8.6 kHz. The dashed curve with open circles shows an input-output function for an 8.6-kHz tone alone, referred to as the probe. The solid curves show input-output functions when a suppressor tone was added to the probe. The suppressor was presented at each of several overall levels, from 60 to 90 dB SPL, as indicated by the key in the figure. The solid curves are truncated at the level where the suppressor led to an increase in response rather than a decrease. Redrawn from Ruggero, Robles and Rich (1992).

tone with frequency $2f_1 - f_2$ produces a local maximum on the BM at the place tuned to $2f_1 - f_2$.

Human listeners can sometimes hear these additional tones. The one with frequency $2f_1 - f_2$ is especially easy to hear. It is audible even for relatively low levels of the primary tones when $f_2$ is about 1.2 times $f_1$ (Smoorenburg, 1972a, 1972b). For example, if the two primary tones have frequencies 1000 and 1200 Hz, then a tone is heard with frequency 800 Hz. The combination tone with frequency $2f_1 + f_2$ is much harder to hear. This may be the case because this tone has a higher frequency than the two primary tones. Although $2f_1 + f_2$ is probably generated at the point where the response patterns of $f_1$ and $f_2$ overlap, it does not propagate to the location on the BM tuned to $2f_1 + f_2$; this would involve propagation in the ‘wrong’ direction, from the apex towards the base.

IV.6 RESPONSES OF THE BASILAR MEMBRANE TO COMPLEX SOUNDS

Consider the response of the BM to two sinusoids, of different frequencies, presented simultaneously. Assume that the sinusoids are equal in level, so that two-tone suppression is small. The kind of pattern of vibration that occurs depends on the frequency separation of the two sinusoids. If this is very large, then the two sinusoids produce two, effectively separate, patterns of vibration on the BM. Each produces a maximum at the place on the BM which would have been excited most had that component been presented alone. Thus, the response of the BM to a low-frequency sinusoid of
moderate intensity is essentially unaffected by a high-frequency sinusoid, and vice versa. In this case, the BM behaves like a frequency analyser, breaking down the complex sound into its sinusoidal components. Correspondingly, when we listen to two sinusoids with widely spaced frequencies, we hear two separate tones, with two different pitches. When the two sinusoids are relatively close in frequency, however, the patterns of vibration on the BM interact, so that some points on the BM respond to both of the sinusoids. At those points, the displacement of the BM as a function of time is not sinusoidal but is a complex waveform resulting from the interference of the two sinusoids. When the two sinusoids are sufficiently close in frequency, there is no longer a separate maximum in the pattern of vibration for each of the component sinusoids; instead there is a single, broader, maximum. Thus, the BM has failed to separate (resolve) the individual frequency components. Correspondingly, when two sinusoids are very closely spaced in frequency, we cannot hear two separate tones, each with its own pitch; rather, we hear a single sound corresponding to the mixture. This is described more fully in Chapter 3, Section VIII.

Consider now the more complex case of the pattern of responses on the BM to a periodic complex tone, such as a voiced vowel or an instrument playing a note. Such a tone typically contains many harmonics of a common fundamental frequency. For example, the note A4 would have a fundamental component with a frequency of 440 Hz, and higher harmonics with frequencies of 880, 1320, 1760 . . . Hz. The harmonics are equally spaced on a linear frequency scale. However, the mapping of CF to distance along the BM roughly follows a logarithmic scale. For example, sinusoids with frequencies of 400, 800, 1600 and 3200 Hz would produce peaks that were roughly equally spaced along the BM (see Figure 1.8 above). When a harmonic complex tone is presented to the ear, the lower harmonics each give rise to a separate peak on the BM, while the higher harmonics give responses that overlap, so that there are not distinct peaks corresponding to the individual harmonics. A perceptual consequence of this is that individual low harmonics can often be ‘heard out’ as separate tones, while higher harmonics cannot be individually heard; this is described more fully in Chapter 3, Section VII. These factors play a crucial role in the perception of complex tones, as is explained in Chapter 6, Section II.

IV.7 OTOACOUSTIC EMISSIONS

Evidence supporting the idea that there are active biological processes influencing cochlear mechanics has come from a remarkable phenomenon first reported by Kemp (1978), although predicted by Gold (1948). If a low-level click is applied to the ear, then it is possible to detect sound being reflected from the ear, using a microphone sealed into the ear canal. The early part of this reflected sound appears to come from the middle ear, but some sound can be detected for delays from 5 to 60 ms following the instant of click presentation. These delays are far too long to be attributed to the middle ear, and they almost certainly result from activity in the cochlea itself. The reflected sounds are known as evoked otoacoustic emissions. They have also been called Kemp echoes and cochlear echoes.
Although the input click in Kemp’s experiment contained energy over a wide range of frequencies, only certain frequencies were present in the reflected sound. Kemp suggested that the reflections are generated at points on the BM, or in the IHC/OHC transduction mechanism, where there is a gradient or discontinuity in the mechanical or electrical properties. The response is nonlinear, in that the reflected sound does not have an intensity in direct proportion to the input intensity. In fact, the relative level of the reflection is greatest at low sound levels; the emission grows about 3 dB for each 10 dB increase in input level. This nonlinear behaviour can be used to distinguish the response arising from the cochlea from the linear middle ear response. Sometimes the amount of energy reflected from the cochlea at a given frequency may exceed that which was present in the input sound (Burns, Keefe and Ling, 1998). Indeed, many ears emit sounds in the absence of any input, and these can be detected in the ear canal (Zurek, 1981). Such sounds are called spontaneous otoacoustic emissions, and their existence indicates that there is a source of energy within the cochlea which is capable of generating sounds. Kemp (2002) and others have suggested that the emissions are a by-product of the active mechanism.

Cochlear emissions can be very stable in a given individual, both in waveform and frequency content, but each ear gives its own characteristic response. Responses tend to be strongest between 500 and 2500 Hz, probably because transmission from the cochlea back through the middle ear is most efficient in this range, as described earlier. Cochlear emissions can be measured for brief tone bursts as well as clicks, and it is even possible to detect a reflected component in response to continuous stimulation with a pure tone.

When the ear is stimulated with two tones, an emission may be detected at the frequency of one or more combination tones, particularly $2f_1 - f_2$. Such emissions are called distortion-product otoacoustic emissions. This confirms that the combination tone is present as a mechanical disturbance in the cochlea, as a travelling wave on the BM.

Sometimes the transient stimulation used to evoke a cochlear echo induces a sustained oscillation at a particular frequency, and the subject may report hearing this oscillation as a tonal sensation. The phenomenon of hearing sound in the absence of external stimulation is known as tinnitus. It appears that tinnitus may arise from abnormal activity at several different points in the auditory system, but in a few cases it corresponds to mechanical activity in the cochlea.

In summary, several types of otoacoustic emissions can be identified, including evoked emissions, spontaneous emissions and distortion-product emissions. While the exact mechanism by which otoacoustic emissions are generated is not understood, there is agreement that it is connected with the active mechanism in the cochlea.

**V NEURAL RESPONSES IN THE NORMAL AUDITORY NERVE**

Most studies of activity in the auditory nerve have used electrodes with very fine tips, known as microelectrodes. These record the nerve impulses, or spikes, in single
auditory nerve fibres (often called single units). The main findings, summarized below, seem to hold for most mammals.

V.1 SPONTANEOUS FIRING RATES AND THRESHOLDS

Most neurones show a certain baseline firing rate, called the spontaneous rate, in the absence of any external stimulus. Liberman (1978) presented evidence that auditory nerve fibres could be classified into three groups on the basis of their spontaneous rates. About 61% of fibres have high spontaneous rates (18–250 spikes per second), 23% have medium rates (0.5–18 spikes per second) and 16% have low spontaneous rates (less than 0.5 spikes per second). The spontaneous rates are correlated with the position and size of the synapses of the neurones on the IHCs. High spontaneous rates are associated with large synapses, primarily located on the side of the IHCs facing the OHCs. Low spontaneous rates are associated with smaller synapses on the opposite side of the IHCs. The spontaneous rates are also correlated with the thresholds of the neurones. The threshold is the lowest sound level at which a change in response of the neurone can be measured. High spontaneous rates tend to be associated with low thresholds and vice versa. The most sensitive neurones may have thresholds close to 0 dB SPL, whereas the least sensitive neurones may have thresholds of 80 dB SPL or more.

V.2 TUNING CURVES AND ISO-RATE CONTOURS

The tuning of a single nerve fibre is often illustrated by plotting the fibre’s threshold as a function of frequency. This curve is known as the tuning curve or frequency-threshold curve (FTC). The stimuli are usually tone bursts, rather than continuous tones. This avoids the effects of long-term adaptation (a decrease in response over time that can occur with continuous stimulation), and also makes it easier to distinguish spontaneous from evoked neural activity. The frequency at which the threshold of the fibre is lowest is called the characteristic frequency (CF) (the same term is used to describe the frequency to which a given place on the BM is most sensitive). Some typical tuning curves are presented in Figure 1.13. On the logarithmic frequency scale used, the tuning curves are usually steeper on the high-frequency side than on the low-frequency side. It is generally assumed that the tuning seen in each single auditory nerve fibre occurs because that fibre responds to activity in a single IHC at a particular point on the BM. Iso-velocity tuning curves on the BM are similar in shape to neural FTCs (Khanna and Leonard, 1982; Sellick, Patuzzi and Johnstone, 1982; Robles, Ruggero and Rich, 1986; Ruggero et al., 1997).

The CFs of single neurones are distributed in an orderly manner in the auditory nerve. Fibres with high CFs are found in the periphery of the nerve bundle, and there is an orderly decrease in CF towards the centre of the nerve bundle (Kiang et al., 1965). This kind of arrangement is known as tonotopic organization and it indicates that the place representation of frequency along the BM is preserved as a place representation in the auditory nerve.
In order to provide a description of the characteristics of single fibres at levels above threshold, iso-rate contours can be plotted. To determine an iso-rate contour, the intensity of sinusoidal stimulation required to produce a predetermined firing rate in the neurone is plotted as a function of frequency. The resulting curves are generally similar in shape to tuning curves, although they sometimes broaden at high sound levels. Also, for high CFs, the frequency at the tip (the lowest point on the curve) may decrease slightly with increases in the predetermined firing rate. This reflects the change in BM tuning with level described earlier.

V.3 RATE-VERSUS-LEVEL FUNCTIONS

Figure 1.14 shows schematically how the rate of discharge for three auditory nerve fibres changes as a function of stimulus level. The curves are called rate-versus-level functions. In each case, the stimulus was a sinusoid at the CF of the neurone. Consider first the curve labelled (a). This curve is typical of what is observed for neurones with high spontaneous firing rates. Above a certain sound level the neurone no longer responds to increases in sound level with an increase in firing rate; the neurone is said to be saturated. The range of sound levels between threshold and the level at which saturation occurs is called the dynamic range. For neurones with high spontaneous rates, this range is often quite small, about 15–30 dB. Curve (b) is typical of what is observed for neurones with medium spontaneous rates. The threshold is slightly higher than for (a) and the dynamic range is slightly wider. Curve (c) is typical of what is observed for neurones with low spontaneous rates. The threshold is higher than for (b). The firing rate at first increases fairly rapidly with the increasing sound level, but
Figure 1.14. Schematic examples of how the discharge rates of single auditory neurones vary as a function of stimulus level. The curves are called rate-versus-level functions. In each case, the stimulus was a sinusoid at the CF of the neurone. Curves (a), (b) and (c) are typical of what is observed for neurones with high, medium and low spontaneous firing rates, respectively.

then the rate of increase slows down. The firing rate continues to increase gradually with increasing sound level over a wide range of levels. This has been called sloping saturation (Sachs and Abbas, 1974; Winter, Robertson and Yates, 1990).

The shapes of rate-versus-level functions can be understood in terms of two functions (Yates, 1990; Patuzzi, 1992). This is illustrated in Figure 1.15. The first function is the input-output function of the BM, illustrated schematically in the top-right panel. The second is the function relating the spike rate in a specific neurone to the magnitude of the BM response. This second function is similar in form for different neurones, showing saturation when the BM amplitude is a certain factor above the value required for threshold, but it varies in the magnitude required for threshold. Three such functions are illustrated schematically in the top-left panel of Figure 1.15. The rate-versus-level functions corresponding to these three functions are shown in the bottom-right panel.

The variation across neurones depends mainly on the type of synapse, as discussed earlier. Neurones with low thresholds have large sensitive synapses. They start to respond at very low sound levels, where the input-output function on the BM is nearly linear. As the sound level increases, the BM displacement increases
Figure 1.15. Schematic illustration of how the shapes of rate-versus-level functions can be accounted for in terms of the BM input-output function (top-right panel), and the functions relating neural firing rate (APs/sec) to amplitude on vibration on the BM (top-left panel). Three such functions are shown, corresponding to synapses with different sensitivities. The resulting three rate-versus-level functions are shown in the bottom panel. Adapted from Patuzzi (1992).

... in a nearly linear manner, and the neurone saturates relatively early, giving a small dynamic range, as shown by the left-most curve in the lower panel. Neurones with higher thresholds have less sensitive synapses. They respond over the range of sound levels where the BM input-output function shows a strong compressive nonlinearity. Hence, a large increase in sound level is needed to increase the BM displacement to the point where the neurone saturates, and the neurone has a wide dynamic range, as shown by the right-most curve in the lower panel.

V.4 TWO-TONE SUPPRESSION

Auditory neurones show an effect that is exactly analogous to the two-tone suppression on the BM that was described earlier. Indeed, the neural effect was discovered long before the BM effect. The tone-driven activity of a single fibre in response to one tone can be suppressed by the presence of a second tone. This was originally called two-tone inhibition (Sachs and Kiang, 1968), although the term two-tone suppression is now generally preferred, since the effect does not appear to involve neural inhibition. Typically the phenomenon is investigated by presenting a tone at, or close to, the CF of a neurone. A second tone is then presented, its frequency and intensity are varied and the effects of this on the response of the neurone are noted. When the frequency and
Figure 1.16. Neurophysiological data from Arthur, Pfeiffer and Suga (1971). The open circles show the tuning curve (threshold versus frequency) of a single neurone with a CF at 8 kHz. The neurone was stimulated with a tone at CF, and just above threshold (indicated by the open triangle). A second tone was then added and its frequency and intensity were varied. Any tone within the shaded areas bounded by the solid circles reduced the response to the tone at CF by 20 % or more. These are the suppression areas.

intensity of the second tone fall within the excitatory area bounded by the tuning curve, this usually produces an increase in firing rate. However, when they fall just outside that area, the response to the first tone is reduced or suppressed. The suppression is greatest when the suppressor falls in one of two frequency regions on either side of the excitatory response area, as illustrated in Figure 1.16. The suppression begins very quickly when the suppressor is turned on, and ceases very quickly when it is turned off (Arthur, Pfeiffer and Suga, 1971). This is consistent with the likely origin of the suppression as a mechanical effect on the BM.

V.5 PHASE LOCKING

In response to a sinusoid with a frequency below about 5 kHz, the nerve firings tend to be phase locked, or synchronized, to the evoked waveform on the BM. A given nerve fibre does not necessarily fire on every cycle of the stimulus but, when firings do occur, they occur at roughly the same phase of the waveform each time. Thus, the time intervals between firings are (approximately) integer multiples of the period of the stimulating waveform. For example, a 500-Hz tone has a period of 2 ms; the waveform repeats regularly every 2 ms. The intervals between nerve firings in response to a 500-Hz tone are approximately 2, or 4, or 6, or 8 ms, and so on. Neurones do not fire in a completely regular manner, so that there are not exactly 500, or 250 or 125 spikes/s. However, information about the period of the stimulating
PHYSIOLOGICAL ASPECTS OF COCHLEAR HEARING LOSS

Figure 1.17. Interspike interval histograms for a single auditory neurone (in the squirrel monkey) with a CF of 1.6 kHz. The frequency of the sinusoidal input and the mean response rate in spikes per second are indicated above each histogram. All tones had a level 80 dB SPL. Notice that the time scales in E and F differ from those in A to D. Redrawn from Rose et al. (1968).

Waveform is carried unambiguously in the temporal pattern of firing of a single neurone.

One way to demonstrate phase locking in a single auditory nerve fibre is to plot a histogram of the time intervals between successive nerve firings. Several such interspike interval histograms for a neurone with a CF of 1.6 kHz are shown in Figure 1.17. For each of the different stimulating frequencies (from 0.408 to 2.3 kHz in this case), the intervals between nerve spikes lie predominantly at integer multiples of the period of the stimulating tone. These intervals are indicated by dots below each abscissa. Thus, although the neurone does not fire on every cycle of the stimulus, the distribution of time intervals between nerve firings depends closely on the frequency of the stimulating waveform.

Phase locking does not occur over the whole range of audible frequencies. In most mammals, it becomes progressively less precise for stimulus frequencies above 1 kHz, and it disappears completely at about 4–5 kHz (Rose et al., 1968), although the exact upper limit varies somewhat across species (Palmer and Russell, 1986). Phase locking improves in precision with increasing sound level at low levels, and then stays roughly constant in precision over a very wide range of sound levels.

Information from phase locking contributes to the ability to localize sounds in space, and it probably also plays a role in the perception of pitch and in the perception...
VI TYPES OF HEARING LOSS

A conductive hearing loss is caused by a reduced efficiency of sound transmission through the outer and/or middle ear. This may be caused by wax (cerumen) in the ear canal, damage to the eardrum produced by infection or trauma, damage to or stiffening of the ossicles in the middle ear, or fluid in the middle ear caused by infection. It results in an attenuation of sound reaching the cochlea, so that sounds appear quieter than normal. The amount of loss may vary with frequency, so sounds may appear to have a somewhat different tonal quality from normal. However, these are the main perceptual consequences of a conductive loss; unlike cochlear hearing loss, it does not generally result in marked distortions or abnormalities in other aspects of sound perception. Conductive hearing loss can often be treated by drugs (to cure infections) or surgery.

Cochlear hearing loss involves damage to the structures inside the cochlea. It can arise in many ways, for example by exposure to intense sounds or to ototoxic chemicals (such as certain antibiotics, drugs used to treat high blood pressure or solvents), by infection, by metabolic disturbances, by some forms of allergies, by autoimmune disorders and as a result of genetic factors. These agents can produce a variety of types of damage to the cochlea, and, to complicate matters further, the damage may extend beyond the cochlea. For example, an infection may produce damage at several sites, such as the auditory nerve and higher centres in the auditory pathway. When both cochlear and neural structures are involved, the more general term sensorineural hearing loss is used.

Finally, hearing loss can occur through damage to structures or neural systems occurring at a level in the auditory system beyond the cochlea, for example in the auditory nerve or the auditory cortex. Such types of hearing loss are given the general name retrocochlear hearing loss. A relatively common cause of retrocochlear loss is the growth of a benign tumour (often called an acoustic neuroma or vestibular schwannoma) which presses on the auditory nerve.

Although sensorineural hearing loss can involve structures other than the cochlea, it is common for the most serious damage to occur within the cochlea. This is probably true for the majority of cases of presbyacusis, the hearing loss that is associated with ageing. Furthermore, it is common for the damage to be largely confined to certain specific structures within the cochlea. This book is concerned with cases where the hearing loss arises primarily from damage to the cochlea. Henceforth, when the phrase ‘hearing impairment’ is used, it should be taken to imply hearing impairment caused by cochlear damage. However, it should be borne in mind that, while it is relatively easy to produce ‘pure’ cochlear damage in animal models, such ‘pure’ damage may be relatively rare in hearing-impaired humans.

In the last twenty years there has been a considerable increase in understanding of the physiology and function of the normal cochlea. Along with this has come an
improved understanding of the changes in function that are associated with cochlear hearing loss. The rest of this chapter reviews the structure and function of the impaired cochlea.

**VII PHYSIOLOGY OF THE DAMAGED COCHLEA**

There is strong evidence that the functioning of the normal cochlea depends upon the operation of an active mechanism that is linked to the integrity of the OHCs. This mechanism may involve feedback of energy onto the BM, via the OHCs, and it plays an important role in producing the high sensitivity of the BM to weak sounds and the sharp tuning on the BM. The normal BM has a strongly nonlinear response, which results in compressive input-output functions, two-tone suppression and combination-tone generation.

Cochlear hearing loss often involves damage to the OHCs and IHCs; the stereocilia may be distorted or destroyed, or entire hair cells may die. The OHCs are generally more vulnerable to damage than the IHCs. Some examples of OHC damage are shown in Figure 1.18. When OHCs are damaged, the active mechanism tends to be reduced in effectiveness or lost altogether. The function of the OHCs can also be adversely affected by malfunctioning of the stria vascularis (Schmiedt, 1996). As a result, several changes occur: the sensitivity to weak sounds is reduced; so sounds need to be more intense to produce a given magnitude of response on the BM, the tuning curves on the BM become much more broadly tuned and all of the frequency-selective nonlinear effects weaken or disappear altogether.

**VII.1 BASILAR MEMBRANE RESPONSES**

There have been many studies showing that the responses of the BM are highly physiologically vulnerable. One example has already been given; see Figure 1.9. Generally, the tuning on the BM becomes less sharp, and the sensitivity around the tip is reduced, when the cochlea is damaged. In the great majority of studies, the changes in BM responses have been associated with some form of damage to the OHCs, either directly or via some form of metabolic disturbance.

The effects of cochlear damage on the input-output functions of the BM of a chinchilla are illustrated in Figure 1.19 (Ruggero and Rich, 1991). The solid curve with black squares, labelled ‘Before’, shows the input-output function obtained when the cochlea was in good condition; the stimulus was a sinusoid with frequency corresponding to the CF of 9000 Hz. The curve shows a compressive nonlinearity for input sound levels between about 30 and 90 dB SPL. In contrast, the response to a sinusoid with a frequency of 1000 Hz, well below the CF, is steeper and is almost linear (solid curve with open circles).

To manipulate the functioning of the cochlea, the animal was injected with furosemide (also known as frusemide), a diuretic that is known to disrupt hair cell potentials, and hence disrupt the function of the hair cells. The dashed curves in
Figure 1.18. Examples of patterns of damage to the OHCs. For the top two panels, the damage was caused by the administration of aminoglycosides. For the bottom panel, the damage was of unknown origin. In the upper panel, the stereocilia of some OHCs are missing. In the middle panel, the stereocilia of the OHCs are mostly either missing or grossly deformed. In the bottom panel, the stereocilia of the OHCs are completely missing over a certain region. The bottom panel also shows expansion of the supporting cells to replace lost hair cells. The electron micrographs were supplied by Dr Andrew Forge of the Institute of Laryngology and Otology, University College London Medical School.
Figure 1.19 were obtained at various times after injection of the drug; the time is indicated by a range in minutes next to each curve. Shortly after the injection (11–19 minutes), the input-output function for the CF tone was markedly altered. The biggest alteration was at low sound levels. To produce a given response on the BM (say, 40 \mu m/s), the input level had to be increased by about 25 dB relative to the level measured before the injection. However, the response to a CF tone at a high level (80 dB SPL) was almost normal. This is consistent with the idea that the contribution of the active mechanism reduces progressively as the sound level is increased above about 40 dB SPL. After a sufficiently long time (112–118 minutes), the input-output function returned to normal. In this case, the cochlear damage was reversible. Larger doses of the drug, or treatment with other drugs, can result in permanent cochlear damage.

Figure 1.19. Input-output functions on the BM immediately preceding (solid lines) and following (broken lines) an intravenous furosemide injection. See text for details. Redrawn from Ruggero and Rich (1991).

VII.2 NEURAL RESPONSES

Some of the first evidence for a physiologically vulnerable active mechanism came from studies of the responses of single neurones in the auditory nerve. Robertson and Manley (1974) showed that the normal, sharp tuning seen in auditory neurones could be altered by reducing the oxygen supply to the animal. The tuning curves
became less sharp, and at the same time the sensitivity around the tip decreased. These changes were similar to those found subsequently in BM responses with similar manipulations (see Figure 1.9 above). The changes in BM tuning and sensitivity found by Robertson and Manley were reversible. Similar effects were reported by Evans (1975), who also found that a reversible degradation in tuning could be produced by the ototoxic agents cyanide and furosemide. Evans and Harrison (1976) used the drug kanamycin to produce selective damage to the OHCs. They found that the threshold and tuning properties of auditory nerve fibres were dependent on the integrity of the OHCs.

VII.3 STRUCTURE–FUNCTION CORRELATION

Liberman and his colleagues used noise exposure and ototoxic drugs, separately or in combination, to produce a variety of types of damage to the hair cells in the cochlea. They then measured neural responses and compared them with structural changes in the cochlea (for a review, see Liberman, Dodds and Learson, 1986). After studying the properties of a given single neurone, the neurone was injected with horseradish peroxidase. This labelled the neurone so that it could be traced to the IHC with which it synapsed in the organ of Corti. In this way, the neural response properties could be directly compared with the structural changes in the hair cells and the immediately surrounding structures in the organ of Corti.

Figure 1.20 shows the situation when there is partial destruction of the OHCs, with intact IHCs. This pattern of damage is typical of that associated with moderate doses of ototoxic drugs, but is less typical of noise exposure. The left part of the figure schematically illustrates the pattern of structural damage. The view is looking down onto the tops of the hair cells; each stereocilium appears as a small dot. The three

Figure 1.20. The left part shows a schematic diagram of an organ of Corti with sub-total loss of OHCs and intact IHCs. The right part shows a normal neural tuning curve (solid) and an abnormal tuning curve (dotted) associated with this kind of damage. Adapted from Liberman, Dodds and Learson (1986).
rows of OHCs appear at the bottom. In the figure, some OHCs are missing or lack stereocilia (see arrow).

The right-hand part of Figure 1.20 shows a typical normal tuning curve (solid curve) and a tuning curve associated with this type of damage (dotted curve). The normal tuning curve shows a sharply tuned ‘tip’ and a broadly tuned ‘tail’. The abnormal tuning curve also appears to have two sections, but the tip is elevated, and the tail is hypersensitive (i.e. thresholds in the region of the tail are lower than normal). The elevated tip may reflect the operation of the active mechanism at reduced effectiveness. The pattern of results suggests that the active mechanism is tuned to a frequency slightly above the resonance frequency of the passive mechanism at that place. This is consistent with the observation that, in a normal ear and at high CFs, the peak response on the BM shifts towards the base of the cochlea with increasing sound level (see Section IV.2).

Figure 1.21 shows the situation when there is a total loss of OHCs with intact IHCs. This pattern of damage is most easily produced with large doses of ototoxic drugs. The bowl-shaped abnormal tuning curve completely lacks the sharp tip because the active mechanism is completely destroyed. The broad tuning of the curve probably depends largely on the passive mechanism.

Figure 1.22 shows the situation where there is severe damage to the stereocilia of the IHCs and the stereocilia of the first row of OHCs (the row closest to the IHCs). The OHC damage is sufficient to eliminate completely the sharply tuned tip of the tuning curve, suggesting that the active mechanism is particularly dependent on the first row of OHCs. In addition, the whole curve is shifted upwards, that is sensitivity is much less than normal. This overall loss of sensitivity (compare the tuning curves in Figures 1.20 and 1.21) can probably be attributed to the IHC damage. According to Liberman et al. (1986), significant damage to the IHC stereocilia is always associated with reduced sensitivity on the tail of the tuning curve.

Liberman et al. did not find any cases of pure IHC damage, without OHC damage. It appears that the OHCs are generally more vulnerable to damage than the IHCs, and so damage to the IHCs is nearly always associated with damage to the OHCs.
Figure 1.22. As Figure 1.20, but for the situation with severe damage to both OHC and IHC stereocilia. Most of the damage to the OHC stereocilia is confined to the first row of hair cells. However, cases approximating pure IHC damage were sometimes found. Figure 1.23 shows an example of a neurone contacting an IHC missing the entire tall row of stereocilia (see arrow) in a region with only modest disarray of OHC stereocilia. The tuning curve shows a tip that is almost normal in shape, but the tip and tail are both shifted upwards by about 40 dB. It appears that the active mechanism was operating, but the transduction mechanism had greatly reduced sensitivity.

These findings can be summarized as follows. The OHCs are responsible for the sharp tips of the tuning curves. When the OHCs are damaged, the sharp tip becomes elevated, or may disappear altogether. This can cause a threshold elevation around the tip of the tuning curve of 40–50 dB. The condition of the OHCs determines the ratio of tip to tail thresholds. Damage to the IHCs causes an overall loss of sensitivity. This is apparent in the tail of the tuning curve, whether or not the OHCs are damaged. Pure OHC damage either leaves the tail unaffected or causes hypersensitivity (lower thresholds) in the tail. When both OHCs and IHCs are damaged, thresholds are usually

Figure 1.23. As Figure 1.20, but for a situation with moderate damage to IHC stereocilia and minimal damage to OHC stereocilia.
greatly elevated, by 80 dB or more, and the tuning curve is broad, without any sign of a sharp tip.

Liberman et al. found that they could ‘account for’ most of the threshold shift in the neurones by evaluating the presence or absence of OHCs and IHCs and the condition of their stereocilia. Furthermore, behavioural thresholds after noise exposure have been found to be correlated with patterns of damage in hair cell stereocilia (Slepecky et al., 1982). It seems likely that, for many types of acquired cochlear hearing loss, the primary cause of the loss is loss of function of the OHCs and/or IHCs. For moderate losses, where thresholds are elevated by less than 50 dB, it may often be the case that the main cause of the loss is damage to the OHCs, with consequent impairment of the active mechanism. In this case, the loss of absolute sensitivity (elevation of threshold) occurs mainly because of reduced responses on the BM to low-level sounds. For more severe losses, it is likely that both OHCs and IHCs are damaged. When the IHCs are damaged, a greater response of the BM is needed to produce a ‘threshold’ amount of neural activity. In extreme cases, the IHCs may be completely non-functional over a certain region of the BM, in which case there is no transduction of BM-vibration in that region and no responses of neurones innervating that region. This is referred to as a dead region and is discussed in more detail in Chapters 2 and 3.

**VII.4 OTOACOUSTIC EMISSIONS**

Evoked otoacoustic emissions are reduced in magnitude by cochlear hearing loss. Human ears with hearing losses exceeding 40–60 dB (see Chapter 2, Section II.4 for the definition of hearing loss) usually show no detectable emissions (Gorga et al., 1997). The emissions appear to be particularly associated with OHC function. The emissions are abolished in ears which have been exposed to intense sounds or to drugs which adversely affect the operation of the cochlea. In the former case, the emissions may return after a period of recovery. This suggests that the emissions are linked to the active mechanism. The measurement of cochlear emissions provides a sensitive way of monitoring the physiological state of the cochlea, and it is now being commonly applied in clinical situations, especially for assessing cochlear function in the very young (Kemp, 2002).

**VII.5 PHASE LOCKING**

The effect of cochlear damage on phase locking is not entirely clear. Harrison and Evans (1979) used the ototoxic drug kanamycin to produce hair cell damage (mainly to OHCs) in the guinea pig, and found that phase locking was not affected. However, Woolf, Ryan and Bone (1981) carried out a similar experiment using the chinchilla and found that phase locking was adversely affected by damage to the OHCs. For neurones with CFs corresponding to frequencies where the behavioural thresholds were elevated by 40 dB or more compared to normal, phase locking was significantly reduced. There was a reduction in the highest frequency at which phase locking could be observed, and the precision of phase locking over the range 0.4–3 kHz was reduced.
The reason for the discrepancy between studies is unclear. It could be related to a species difference. Another possibility is that the data of Harrison and Evans are atypical. Woolf et al. pointed out that the upper limit of phase locking reported by Harrison and Evans (5–7 kHz) was higher than usually observed in the guinea pig. The reason why phase locking should deteriorate as a consequence of cochlear damage is also unclear. Woolf et al. suggested that it could be connected with poorer mechanical coupling between the tallest stereocilia of the OHCs and the tectorial membrane. Whatever the reason, it seems clear that phase locking can sometimes be affected by cochlear damage, and this may have important perceptual consequences.

Cochlear damage can certainly affect phase locking to complex sounds, such as speech. Vowel sounds contain peaks in their spectral envelope at certain frequencies called formants; these correspond to resonances in the vocal tract (see Chapter 8 and Figure 8.3). Each formant is actually defined by several harmonics, whose amplitudes exceed those of adjacent harmonics. The formants are numbered, the lowest in frequency being called F1, the next F2 and so on. The frequencies of the formants, and especially the first two formants, are believed to be important for determining the identities of vowel sounds. At low levels in a normal auditory system, each neurone shows phase locking to a single harmonic or a small group of harmonics whose frequencies lie close to the CF of the neurone. Hence, the temporal response patterns vary markedly across neurones with different CFs. However, at higher levels, the temporal response patterns show a ‘capture’ phenomenon, in which the first two formant frequencies dominate the responses; neurones with CFs that are somewhat removed from a formant frequency may nevertheless show strong phase locking to that formant (Miller et al., 1997). Most of the neurones with mid-range CFs show phase locking either to F1 or to F2. This may partly depend upon a suppression effect, related to two tone suppression, whereby the relatively strong harmonics close to the formant frequencies suppress the responses to weaker harmonics at adjacent frequencies. Whatever the mechanism, the temporal information coded in the phase locking may be used by the auditory system to determine the formant frequencies.

Miller et al. (1997) studied the effect on this phenomenon of cochlear damage caused by exposure to intense sound. After the acoustic trauma, capture by the second formant (which fell in the region of threshold elevation) was not observed; neurones with CFs adjacent to F2 did not show clear phase locking to F2, but showed more complex response patterns. The phase locking to formant frequencies observed in the normal auditory nerve may play an important role in the coding of the formant frequencies in the auditory system. If so, the reduced phase locking associated with cochlear damage might contribute to problems in understanding speech.

VIII CONCLUSIONS

The functioning of the normal cochlea is strongly dependent on an active mechanism that is physiologically vulnerable. This mechanism depends upon the integrity of the OHCs, and particularly their stereocilia. The active mechanism is responsible
for the high sensitivity and sharp tuning of the BM. It is also responsible for a variety of nonlinear effects that can be observed in BM responses and neural responses. These effects include: the nonlinear input-output functions on the BM, the reduction in sharpness of tuning with increasing sound level and combination tone generation. Finally, the active mechanism is probably responsible for the generation of evoked and spontaneous otoacoustic emissions. The active mechanism strongly influences responses on the BM at low and medium sound levels, but its contribution progressively reduces as the sound level increases.

The OHCs are easily damaged by noise exposure, ototoxic chemicals, infection and metabolic disturbances. When they are damaged, the active mechanism is reduced in effectiveness or destroyed completely. This has several important consequences:

1. Sensitivity is reduced, so that the tips of tuning curves are elevated by up to 40–50 dB.
2. The sharpness of tuning on the BM is greatly reduced. The tip of the tuning curve may be elevated or may disappear altogether, leaving only the broad tuning of the passive BM.
3. Nonlinear effects such as compressive input-output functions on the BM, two-tone suppression and combination tone generation are reduced or disappear altogether.
4. Evoked and spontaneous otoacoustic emissions are reduced or disappear, at least in the frequency range corresponding to the damaged place.

The IHCs are the transducers of the cochlea converting the mechanical vibrations on the BM into neural activity. They are less susceptible to damage than the OHCs. When they are damaged, sensitivity is reduced. Damage primarily to the IHCs, with intact OHCs, is rare. When it does occur, sharp tuning may be preserved, but the whole tuning curve is elevated (less sensitive). More commonly, damage occurs both to the OHCs and the IHCs. In this case, the whole tuning curve is elevated, with a greater elevation around the tip than around the tail.

Damage to the hair cells sometimes, but not always, results in a reduction in phase locking; the precision with which neural impulses are synchronized to the cochlear-filtered stimulating waveform may be reduced. The reason why this occurs is unclear, but it may have important perceptual consequences.