7

Physiology and Anatomy of Hearing

The purpose of hearing is to capture acoustic vibrations arriving at the ear and analyse the content of the signal to deliver information about the acoustic surroundings to the higher levels in the brain. The auditory system is the sensory system for the sense of hearing. According to current knowledge, the auditory system divides broadband ear-canal signals into frequency bands, and then conducts a sophisticated analysis on the bands in parallel and in sequence. The human auditory system largely resembles the hearing system in other mammals, but humans have one property that has developed much farther: the ability to analyse and recognize spoken language. As a consequence, the sensitivity to and resolution of some speech- and voice-related features of sound are very good. Thus, a major part of this book is devoted to discussing the different roles of hearing and the auditory perception in typical human communication.

The functional properties, or the *physiology of hearing*, are interesting in the context of understanding communication and engineering applications, especially from a basic research point of view. This topic includes the acoustic-to-mechanic and then to neural conversion occurring in the auditory periphery and the neural functions of the auditory pathway. From a communication point of view, however, a very detailed understanding of the physiology of hearing is not necessary. The *anatomical* structure of hearing is also somewhat interesting, though not very important as such, except in some special cases, such as audiology or spatial hearing. Thus, a brief introduction to both the anatomy and physiology of the auditory system is considered sufficient in this chapter.

7.1   Global Structure of the Ear

Humans, as well as most animals, have two sensors for sound – *the left and right ear* – and a complex neural system to analyse the sound signals received by them. The ear, more specifically the *peripheral auditory system*, consists of the *external ear* for capturing sound waves travelling in the air, the *middle ear* for mechanical conduction of the vibrations, and the *inner ear* for mechanical-to-neural transduction. Neural signals from the periphery are transmitted through the *auditory pathway*, where the neural signals are processed by different nuclei up to the auditory cortex where high-level analysis occurs.
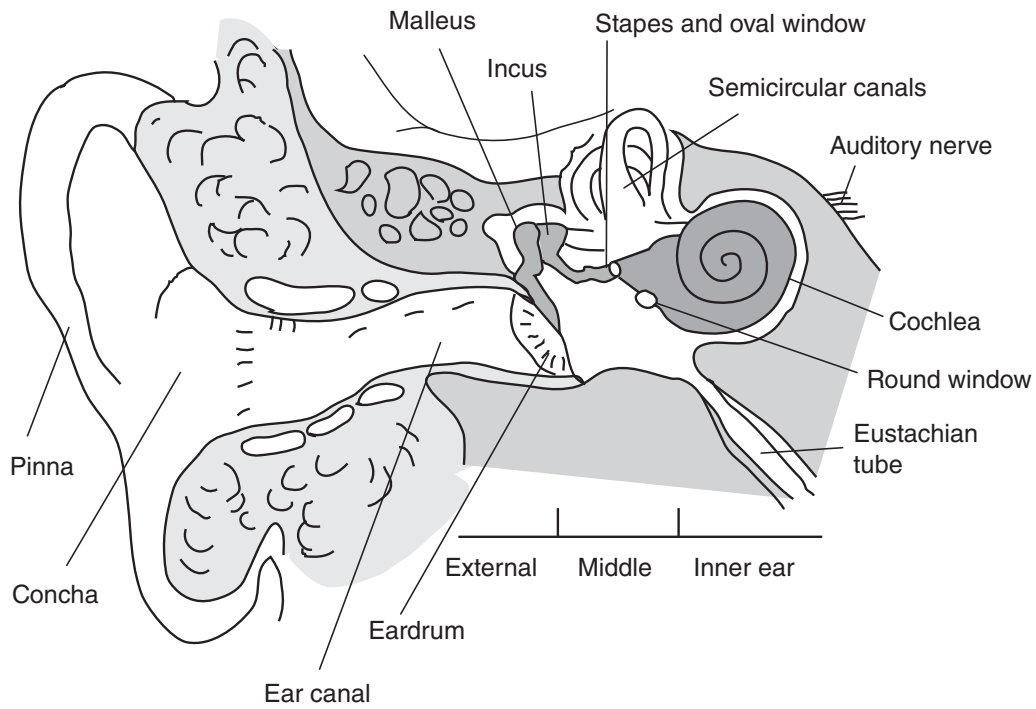
Figure 7.1   Cross-sectional diagram of one ear, showing the external, middle, and inner ear.
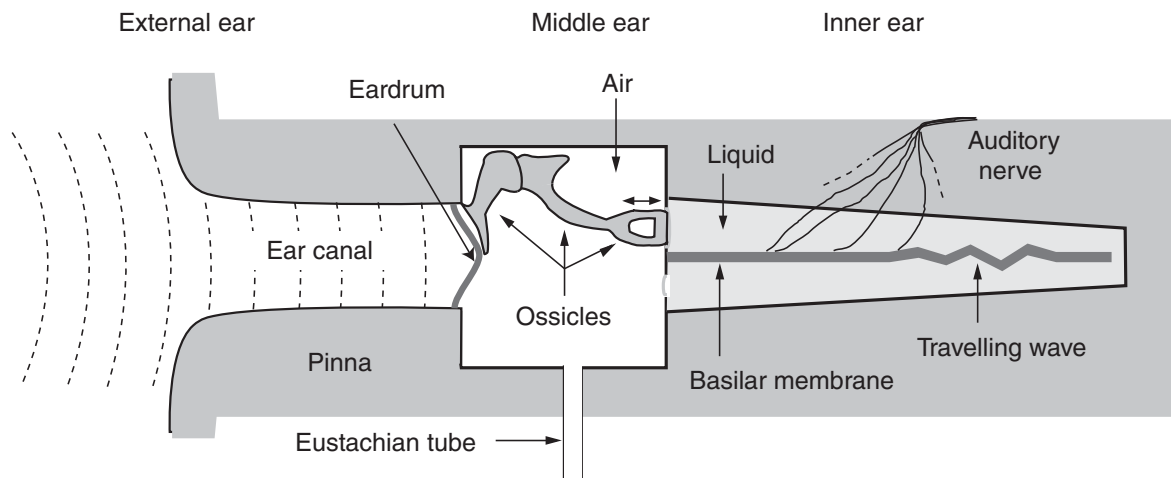

Figure 7.2   A simplified diagram of the ear.

Figure 7.1 depicts a cross-section of one ear, including the external (outer), the middle ear, and part of the inner ear. Figure 7.2 shows a more schematic diagram characterizing the most essential parts of the system and the path from acoustic wave to neural signal.

7.2   External Ear

The external ear (outer ear) consists of the *pinna* with the *concha*, the *ear canal* or *meatus*, and the *eardrum* or *tympanic membrane* as a borderline with the middle ear (Figure 7.1). The external ear is passive and linear, and its functioning is entirely based on the laws of acoustic
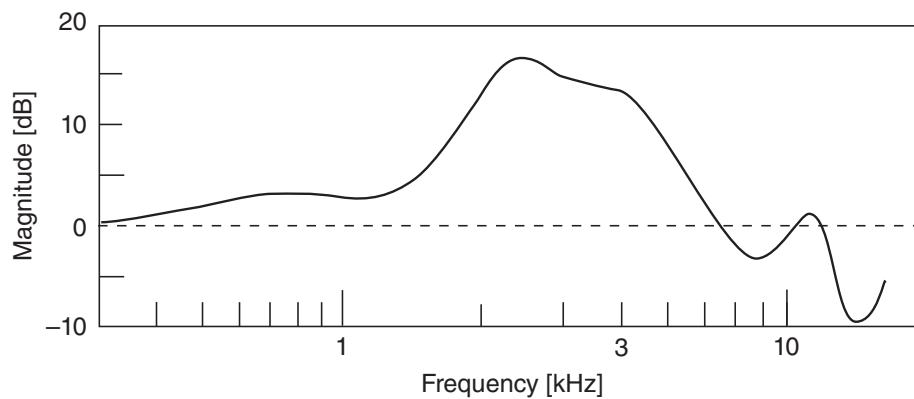
Figure 7.3   An approximate magnitude response of the outer ear at the eardrum to a frontal sound source in free-field conditions.

wave propagation. Passivity means that the external ear does not generate sound energy or 'react' to the sound, it only carries sound waves properly to the eardrum and the middle ear.

Acoustically, the entire head (and shoulders) makes part of the external ear. The distance between the two ears, located at opposite sides of the head, causes an arrival time difference of the sound wavefront that depends on the angle of incidence. A difference in sound level appears at high frequencies due to shadowing by the head to the side that is opposite to the sound source. These phenomena and related effects in spatial hearing are discussed in Chapter 12. In the present chapter, we concentrate on *monaural* phenomena; that is, hearing where interaural differences do not play a role.

The role of the *pinna* is to influence sound propagation to the ear at high frequencies by creating asymmetry, both front–back and top–down, that helps directional hearing. The cavity leading to the ear canal is called the *concha*. Together, the head and the external ear, up to the eardrum, emphasize frequencies of 1–5 kHz considerably, as depicted in Figure 7.3. This is measured for a sound source in front of a listener in free-field conditions. The dependency on the incidence angle is discussed in Chapter 12 based on the concept of the *head-related transfer function* (HRTF).

The *ear canal* (*external auditory meatus*) is a relatively hard-walled tube that is approximately 22.5 mm long and has a diameter of about 7.5 mm, so that its volume is about 1 cm^3. Acoustically, it is a short transmission line where sound waves propagate from the external environment to the eardrum. Because one end is open (low acoustic impedance) and the impedance of the eardrum is higher than for the tube itself, the ear canal acts as a quarter-wavelength resonator, emphasizing signals around frequencies of 3–4 kHz by about 10 dB, attenuating them at around 7–8 kHz, and showing the next (weak) resonance above 10 kHz.

The *eardrum* or the *tympanic membrane* is a membrane that converts sound waves arriving in the air through the ear canal into mechanical vibrations and passes these vibrations to the middle ear.

7.3   Middle Ear

The middle ear, which is located in a small, air-filled cavity between the eardrum and the inner ear, as shown in Figure 7.1, is a transmission system of mechanical vibration from the eardrum through little bones called *ossicles* to the *oval window* that leads to the cochlea of the inner ear. When the eardrum vibrates due to sound entering through the ear canal, the eardrum and
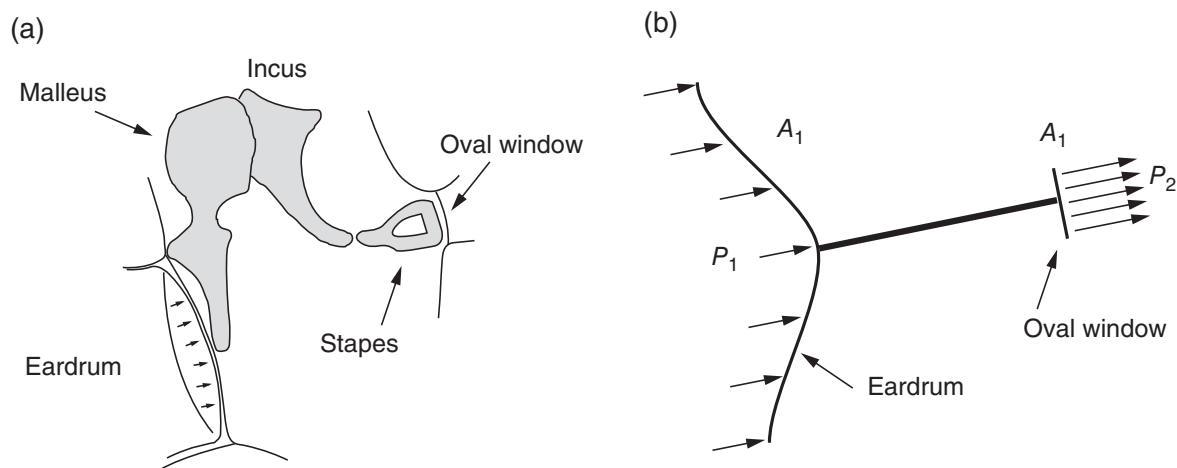
Figure 7.4 The ossicles in the middle ear: (a) the real structure, and (b) impedance matching due to the area ratio.

the ossicles vibrate along and transmit the vibration to the liquid medium in the inner ear. The ossicles consist of three very small bones called the *hammer* or *malleus*, the *anvil* or *incus*, and the *stirrup* or *stapes* (see Figure 7.4a).

The relatively complex construction of the middle ear is a result of evolution to provide sufficiently good impedance matching between different physical media. Since the characteristic impedances of air and water (in the cochlea) are very different (a ratio of about 1:3000), the vast majority of sound energy entering the ear canal would reflect back and only a minimal fraction could proceed to the inner ear (see Equations (2.25)–(2.27) and (2.29) for reflection, transmission, and absorption coefficients) if there were no special mechanism between the air in the ear canal and the liquid in the cochlea. A minor improvement is achieved by the chain of ossicles that works as a lever mechanism. Another, far more efficient effect is obtained by the area ratio of the eardrum and the oval window, as illustrated in Figure 7.4b, that works as an impedance transformer. This mechanism transforms a small pressure with large velocity in the air into a large pressure with small velocity in the liquid of the inner ear. Good efficiency of sound energy transfer is important in order to obtain the extremely high sensitivity that the auditory system has. The motion displacement of the eardrum at middle frequencies (1–4 kHz) at the threshold of hearing is only about 10^{-9} cm, which means just about a tenth of the diameter of a hydrogen atom!

The impedance-matching mechanism improves the pressure transfer by a factor of about 30; that is, by about 30 dB. The transfer function of the middle ear is a band-pass filter where mid-frequencies are emphasized, and the response rolls off at higher and lower frequencies, as characterized in Figure 7.5.

There is a narrow channel called the *Eustachian tube* connecting the middle ear to the oral cavity. The role of this tube is to balance the air pressure between the middle ear and the environment, for example during changes in altitude, such as in aeroplanes during takeoff or landing. Too high or too low a pressure in the middle ear displaces the eardrum and reduces the sensitivity of hearing, or may even cause pain.

Another detail of the middle ear that is related to the auditory function is the *acoustic reflex*. Loud sounds, approximately above 50–60 dB, trigger small muscles to contract so that conduction of sound by the ossicles is reduced (Geisler, 1998). This is understood as a mechanism
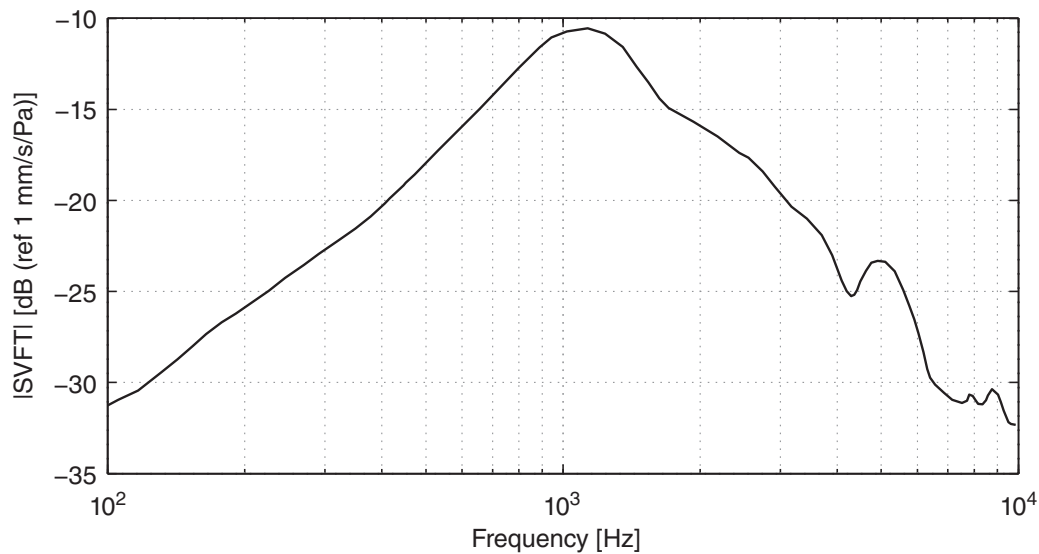
Figure 7.5   The middle ear pressure transfer function from the pressure at the eardrum to the velocity of the stapes window. Adapted from Aibara *et al.* (2001).

against overly loud sounds in order to protect the delicate mechanisms of the inner ear. This protection is, however, not very efficient, since it is slow (with a latency of tens or hundreds of milliseconds) against impulsive sounds and it attenuates prominently only low frequencies, 10–20 dB below 500 Hz and less above 2 kHz (Møller, 2006). However, the system may be effective against some impulsive sounds that the subjects generate themselves. It has been found that the reflex is activated subconsciously if a considerable vibration or sound is imminent, say when running on a hard surface, shouting, or using firearms. This reflex, when on, changes the acoustic impedance of the eardrum, and it can therefore be registered indirectly from the outer ear (Geisler, 1998).

7.4   Inner Ear

The inner ear consists of the cochlea and the semicircular canals. While the latter is a balance sensing organ and has no effect on monaural hearing, the cochlea is a marvellously complex system for transforming mechanically conducted sound from the middle ear first into the vibrations of the basilar membrane and then into neural impulses for higher-level analysis.

7.4.1   *Structure of the Cochlea*

The *cochlea* is a spiral-shaped and liquid-filled tube of about 2.7 turns with a length of 35 mm (see Figure 7.1) (Plack, 2013). The cochlea is in vibroacoustic connection with the middle ear via two ports called the *oval window* and the *round window*. The oval window connects the vibrations from the stapes to the liquid medium inside the cochlea. Mechanically, the most important part inside the cochlea is the *basilar membrane*, which is located between bony shelves along the length of the cochlea, as illustrated in the linearized model of Figure 7.6. There is an opening called the *helicotrema* at the apical end of the basilar membrane, connecting the liquid media on both sides of the membrane. The width of the membrane reaches about 0.5 mm at its widest point, which is at the apical end (Plack, 2013).
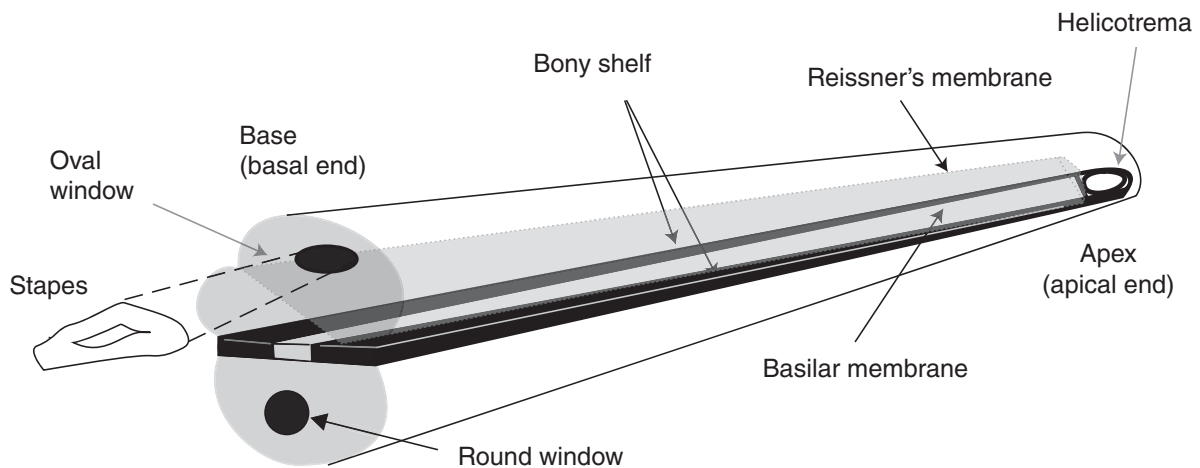
Figure 7.6   The structure of the cochlea depicted as a linear tube instead of its true, spiral form.
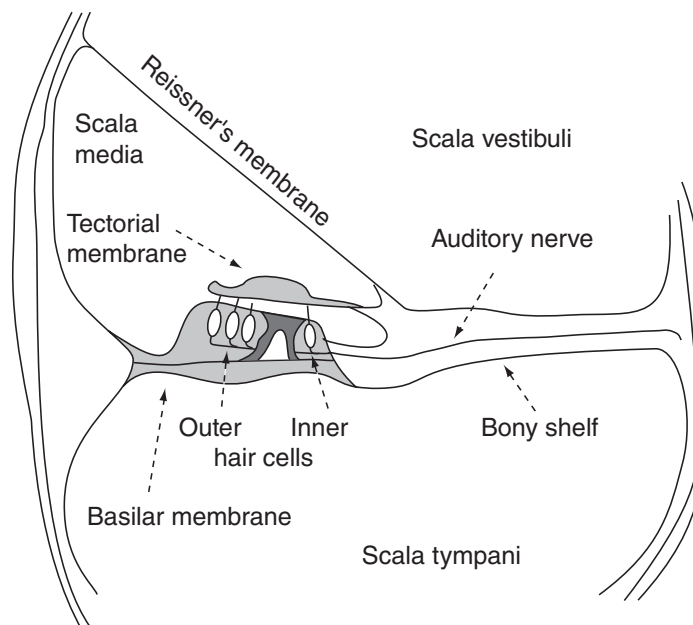
Figure 7.7   Cross-sectional structure of the cochlea.

A cross-section of the cochlea is depicted in Figure 7.7, showing other structural elements inside the tube. The basilar membrane and the bony shelf divide the tube into two halves. The liquid section above the basilar membrane is the *scala vestibuli* and that below the membrane the *scala tympani*. A thin membrane, *Reissner's membrane*, separates one more liquid section called the *scala media*. The scala media has a structure in its bony wall that leaks potassium ions (K^+) into it. Thus, since the ion concentrations of liquids in the scala media and the scala tympani are different, a small electric potential difference exists across the basilar membrane.

A vibration-sensitive structure called the *organ of Corti* resides on the basilar membrane (shaded in Figure 7.7). It has receptors called *hair cells*, which are specialized cells involved in the process of converting the vibration pattern of the basilar membrane into neural impulses in the *auditory nerve* fibres. Hair cells are organized in several rows: one row of *inner hair*
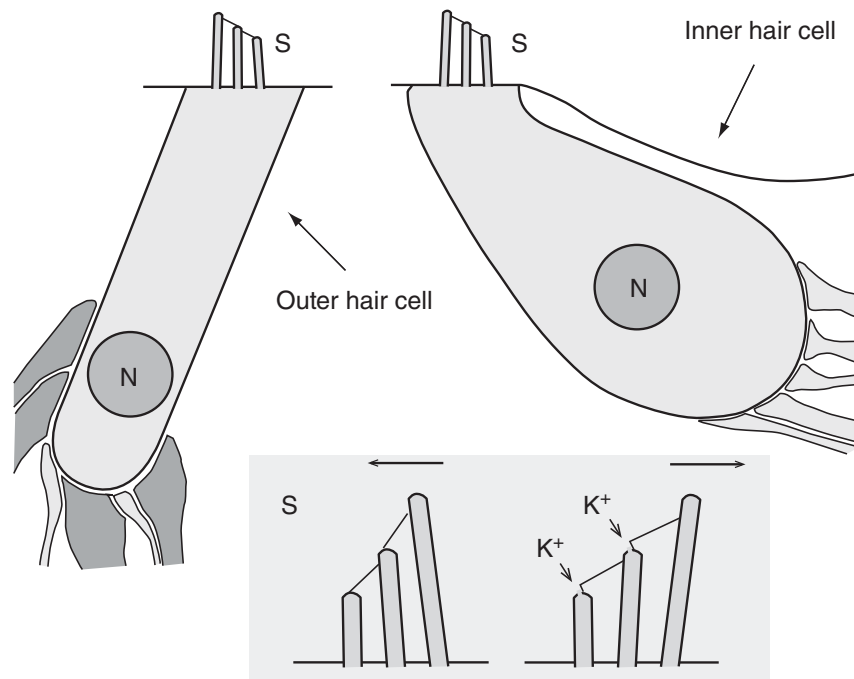
Figure 7.8 Top: A simplified schematic structure of the inner and outer hair cells. N = nucleus and S = stereocilia. Nerve endings at the bottom of a cell form the ascending path if shaded light grey and the descending path if dark grey. Below: Bending the stereocilia right opens the channels for K^+ ions on the tops of the hair cells.

cells and three to five rows of *outer hair cells*. There are about 3500 inner and 12 000 outer hair cells uniformly distributed along the organ of Corti. Each hair cell is equipped with a bunch of hair-like filaments on top of it, called *stereocilia*, structured in a U, V, or W shape. Figure 7.8 characterizes the structure of an inner and an outer hair cell. The lengths of the outer hair cells are about 12 μm at the basal end and about 90 μm at the apical end. Their positioning along the organ of Corti also varies systematically.

There exists one more membrane, the *tectorial membrane*, a gelatinous structure right above the hair cells and stereocilia. When the basilar membrane vibrates, the stereocilia are set into bending motion, and the hair cells are extremely sensitive in reacting to these movements. The inner hair cells are the primary receptors, while the outer hair cells have an active function to control the mechanical vibration state of the system, as discussed below.

7.4.2 Passive Cochlear Processing

The main role of cochlear processing is to act as a frequency-to-place mapping of signal components. Let us first assume that the basilar membrane is passive and time-invariant. The membrane is structured so that it is narrower, less massive, and stiffer at the basal end (oval window side) and gradually changes its characteristics to wider, more massive, and loosely moving at the opposite (apical) end. Notice that the widening of the basilar membrane towards its apical end is in contrast to the narrowing of the cochlear tube (see Figure 7.6). An inherent property of such a system is that each point along the line has a frequency of highest vibratory amplitude; that is, each point on the basilar membrane resonates at a specific frequency, which is called the *best frequency*, or *characteristic frequency*. This phenomenon was found

and studied by von Békésy (1960), earning him the Nobel Prize in Physiology or Medicine in 1961 for his work.

The vibration enters the liquid through the oval window and travels at a very high speed in the duct. The round window acts as a sink for the vibration, and one might think that the vibration finds a route there by penetrating through the basilar membrane. This vibrating liquid column can be seen as a mass–spring system, and, as shown by Equation (2.6), the lower the mass, the higher the resonance frequency. High frequencies thus favour a shorter route, since the mass of the vibrating liquid decreases if the route is shorter. In contrast, the lowest frequencies favour a longer route, since the mass of the system is higher there. Added to this, the vibration favours penetrating the basilar membrane at a position where its frequency matches the characteristic frequency of the membrane. Thus, many different characteristics are optimized to make the frequency-to-place conversion of acoustic vibrations. This conversion is shown schematically in Figure 7.9.

The vibration seen in the basilar membrane has a certain behaviour with space and time, it seems to 'travel' towards the apex, as shown in Figure 7.10. It also has a clear resonance peak. The vibration of the liquid on the basilar membrane is often considered to cause a *travelling wave* to the membrane, which then carries the signal content with it and excites the resonances on the membrane (Yates, 1995). In another explanation, it is assumed that the vibratory
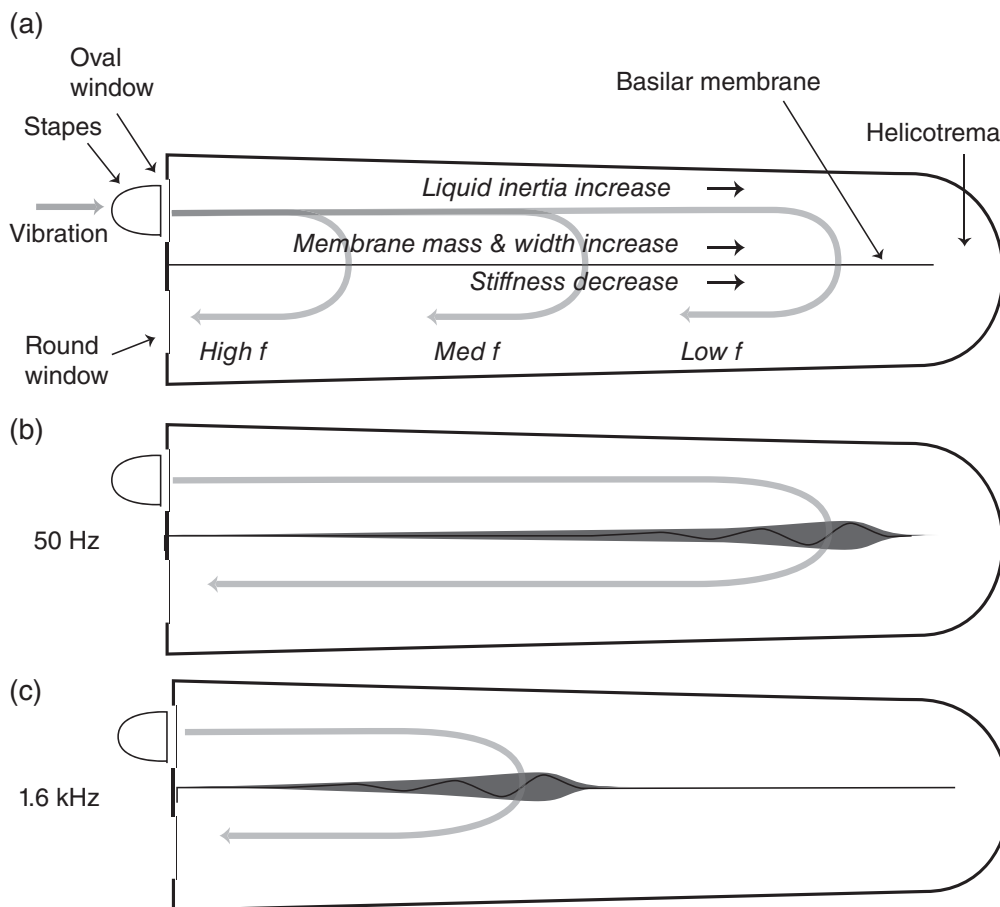
Figure 7.9 (a) A schematic illustration of the acoustic properties of the cochlear duct that change with distance from the stapes. The position at which a component of vibration entering the cochlea through the stapes penetrates the basilar membrane depends on its frequency. (b) and (c) The amplitude envelopes of the travelling waves on the basilar membrane caused by sinusoidal stimuli.
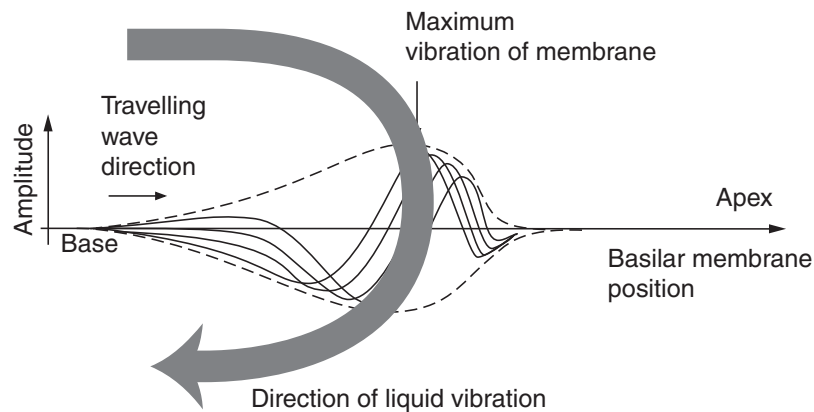
Figure 7.10   A travelling wave on the basilar membrane caused by a vibration in the liquid with the peak amplitude forming at a specific position for a specific stimulus frequency. The solid curves are waveforms at consecutive moments in time, while the dashed curve is the amplitude envelope of the vibration.

excitation from the liquid reaches each position on the membrane virtually at the same time, and that the formation of the wave is caused by the phase response characteristics of different positions of the basilar membrane. It is assumed that the positions with lower characteristic frequency respond a bit later, and such a wave phenomenon occurs (Schnupp *et al.*, 2011).

Figure 7.9 plots the amplitude envelope of a vibration on the basilar membrane for different frequencies. The lowest frequency signals resonate at the end of the membrane, while higher frequencies resonate earlier and then decay rapidly. The highest frequencies (near 20 kHz) have their maximum vibration at the very beginning of the membrane (not shown in Figure 7.9).

7.4.3   Active Function of the Cochlea

The explanation of the basilar membrane and the hair cell function above is qualitatively correct as a first approximation. Georg von Békésy and followers first measured the vibratory patterns of the membrane using ears that were not in prime physiological condition. They found that the frequency selectivity of mechanical vibration was far poorer than the selectivity measured from auditory nerve responses. This was explained by a hypothetical 'second filter' somewhere between these points.

Later studies have found that the collaboration of the basilar membrane and the hair cells makes a complex signal processing system – an *active* cochlea – where neural activity can control the vibratory pattern of the basilar membrane. The hair cells not only sense vibration and code this information to the auditory nerve, but they, particularly the outer hair cells, can actively amplify the mechanical motion. The term *cochlear amplifier* is used to refer to this phenomenon. The gain of the amplifier is shown in Figure 7.11a, where the velocity of the basilar membrane is measured at a single point in the cochlea of a cat, and the ratio between the velocity of the membrane and the pressure of the input signal is plotted as a function of input frequency. The gain function is high and has a high Q (good frequency selectivity; see Equation (3.30) on page 58) for low signal levels at frequencies near 9 kHz (Ruggero *et al.*, 1997). When the input sound pressure level is higher, the best frequency is somewhat lower, and the Q is significantly lower.

The active role of the cochlea and hair cells is based on non-linear positive feedback, whereby very weak signals are strongly emphasized in amplitude and the Q value of the cochlear resonance is improved. The outer hair cells have been shown to be able to modulate their length
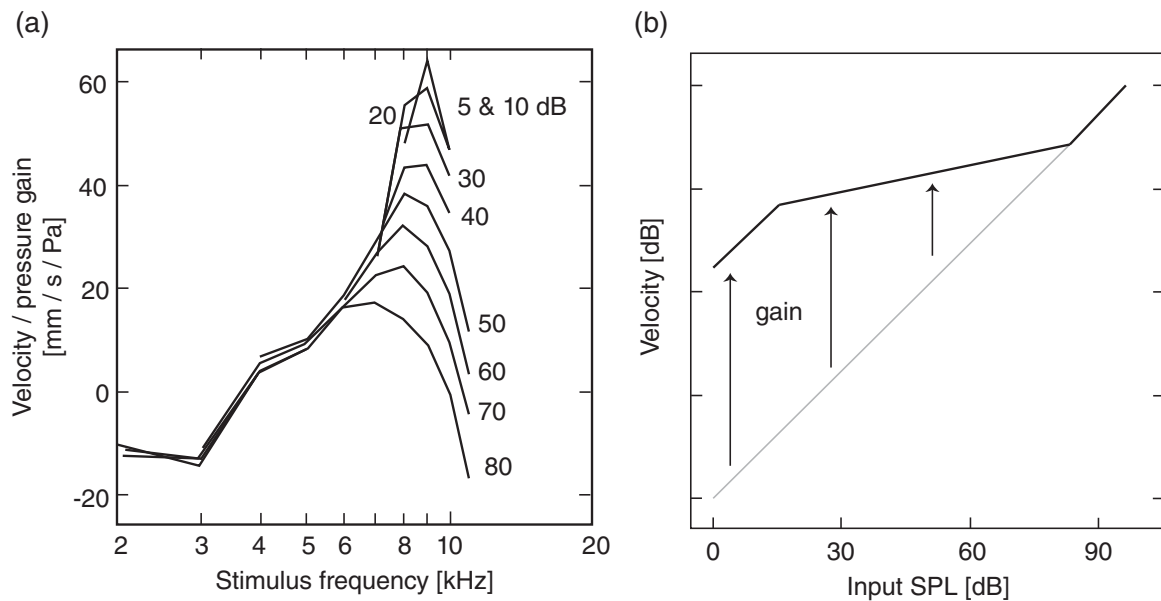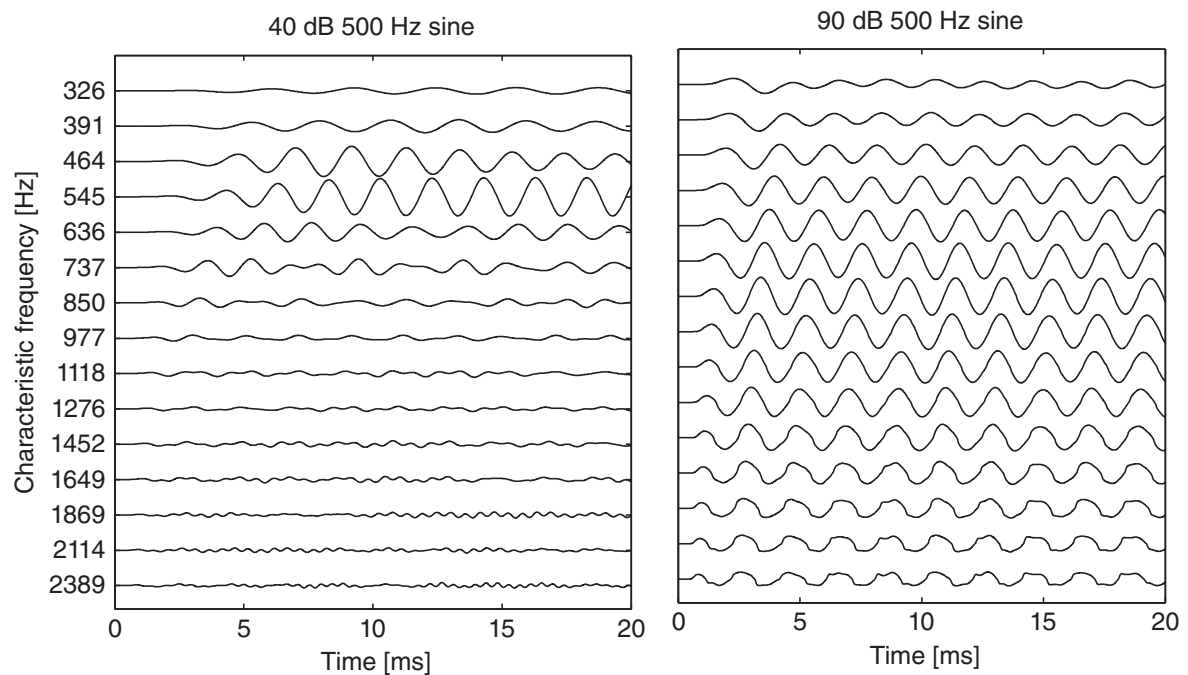
Figure 7.11 (a) The velocity of the basilar membrane recorded at a single point when excited with sinusoids of different frequencies and levels. The gain (velocity of the membrane divided by stimulus pressure) of the active cochlea is shown as a function of stimulus frequency. After Ruggero *et al.* (1997), with permission from The Acoustical Society of America. (b) A schematic plot of the gain function at a single position in the cochlea showing the dependence on the stimulus level.

when they are excited mechanically or by the descending neurons (Schnupp *et al.*, 2011). However, the exact mechanism of *how* the outer hair cells affect the cochlea movement is poorly understood and is not discussed here. A consequence of the active neural involvement in the cochlea is *otoacoustic emissions*, which are discussed more in Section 7.5.

When the stimulus level increases, active boosting will decrease until, for loud speech levels, the signal is no longer amplified. Cochlear compression is characterized in Figure 7.11b as a compressive input–output relation. A consequence of the outer hair cell contribution to the basilar membrane movement is that the system cannot be seen as a simple linear filter bank. It has many non-linear features in responses within frequency, time, and level.

To illustrate this phenomenon more clearly, Figure 7.12 shows the velocity patterns that arise at different positions of the membrane when two 500-Hz sinusoids with levels 40 dB SPL and 90 dB SPL are fed to a non-linear model of the cochlea. The high Q value at low levels of excitation results in considerable activation only near 500 Hz in the 40-dB case. When the SPL of the sinusoid is raised to 90 dB, the activation spreads over a large range towards higher frequencies. This behaviour matches the results shown in Figure 7.11a, where the position on the basilar membrane is sensitive to frequency very selectively only when the excitation level is low, and when the level is increased, it also responds to frequencies *lower* than its best frequency. Notice also in Figure 7.12 that for the 40-dB case the maximum velocity is obtained after 10 ms of stimulus, whereas in the 90-dB case it is reached at 4 ms. This is a result of the active amplification of the vibration, which takes some time to build up.

To show the functioning of the cochlea with time, the response of a non-linear cochlea model (Verhulst *et al.*, 2012) to an impulse train with a repetition rate of 10 ms is shown in Figure 7.13. The membrane responds at high frequencies after each impulse by damping the oscillation like a resonator. At the lowest frequencies the response follows the fundamental frequency of

Figure 7.12 The normalized basilar membrane velocity response to a 500-Hz sinusoidal input a) at 40 dB SPL and b) at 90 dB SPL. Courtesy of Alessandro Altoè.
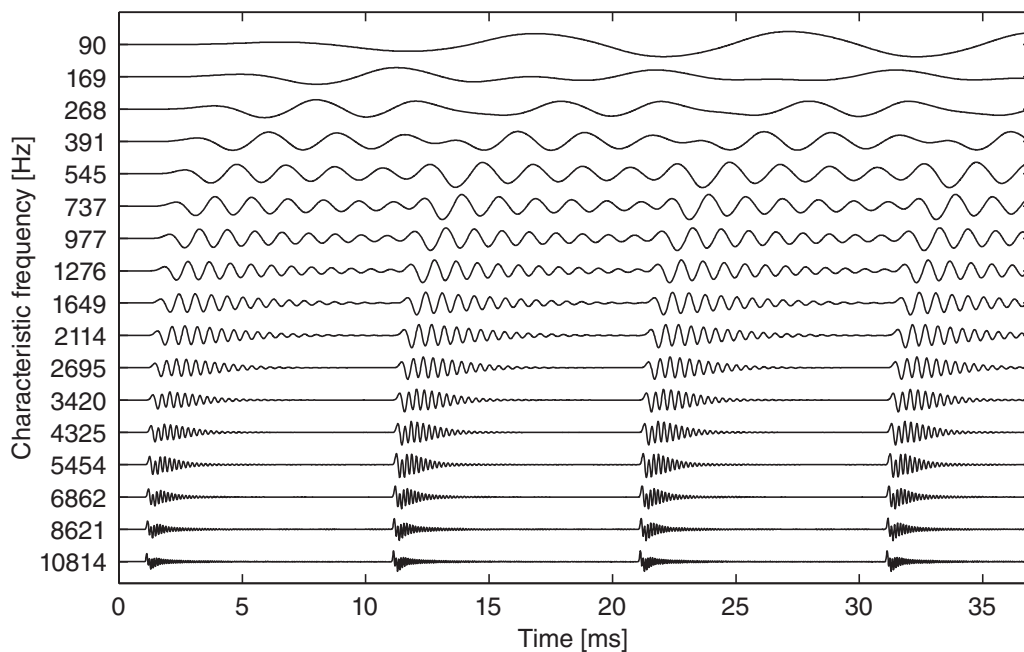
Figure 7.13 The velocity response of the basilar membrane to an impulse train with a repetition rate of 100 Hz as the excitation. The response is shown for different positions on the membrane corresponding to the indicated characteristic frequencies computed using a non-linear model of the cochlea (Verhulst *et al.*, 2012). Courtesy of Alessandro Altoè.

the excitation signal. A noticeable property is the delay of the response that increases towards lower frequencies due to the slower movement of the membrane at the apical end. The response resembles a travelling wave with a velocity that decreases with time.

7.4.4 The Inner Hair Cells

The conversion of the vibrations of the cochlea into the activity of neurons is performed by the inner hair cells. The vibration of the basilar membrane leads to the bending of the stereocilia which, in turn, causes the channels on the tops of the hair cells to open for K^+ ions, as shown in Figure 7.8. The movement of the K^+ ions modulates the potential difference across the membrane of the cell. The changes in the membrane potential trigger the release of *neurotransmitters* at *synaptic junctions* between the inner hair cells and neurons of the auditory nerve – the neurons that make up the auditory nerve are called the *spiral ganglion neurons*. Each inner hair cell is connected to about 20 spiral ganglion neurons, each of which receives an input from one inner hair cell only.

The frequency-dependent behaviour of the basilar membrane means that it acts, together with the signal detection by the hair cells, as a mechanical band-pass filter. A vibratory pattern on the basilar membrane is mediated by hair cells to the auditory nerve, thus coding frequency to place and then to neural fibre position. This position ordering of frequencies from high to low is called *tonotopical organization* (tonotopy), and the ordering is preserved throughout the neural processing stages up to the auditory cortex.

7.4.5 Cochlear Non-Linearities

There are other deviations from linear and time-invariant behaviour found in the neural responses in addition to the compression of dynamics shown in Figure 7.11. The two most prominent *cochlear non-linearities* are called two-tone suppression and combination tones. *Two-tone suppression* is the phenomenon where the addition of a second input tone at another frequency suppresses the activation caused by the test tone at its characteristic frequency (Delgutte, 1990). Figure 7.14 illustrates this effect in the form of a tuning curve where
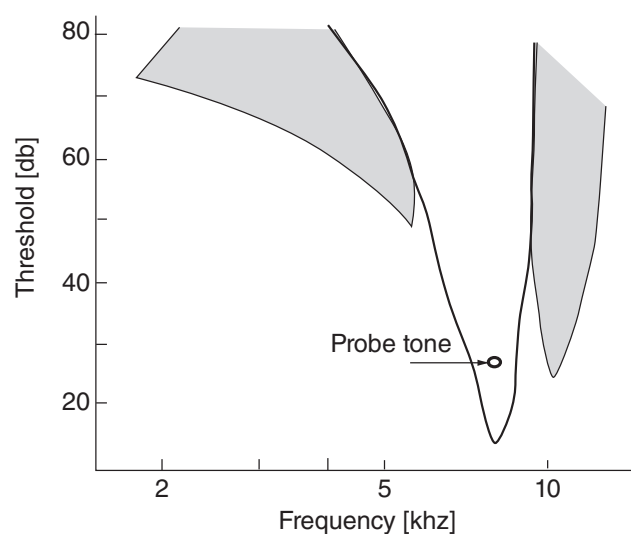
Figure 7.14 Two-tone suppression. A secondary tone with a frequency and level falling in the shaded area will suppress the response to the probe tone at its CF. Adapted from Arthur *et al.* (1971).

the shaded areas represent suppressed behaviour, in which if the second tone falls in these frequency–level areas, the neural activity due to the test tone at its CF will decrease. This phenomenon can be interpreted as a kind of masking, a means for stronger (and potentially more important) signal components to dominate over neighbouring weaker signal components that may be noise or other interfering sounds.

Another non-linearity of common interest is the generation of *combination tones*, a phenomenon common to all non-linear systems. Active collaboration of the basilar membrane and the outer hair cells generates harmonics for a single sinusoidal input, but due to the strong low-pass filtering along the basilar membrane beyond the point of the CF these harmonics do not play a prominent role. Combination tones of two (or more) sinusoids, however, are more audible. The difference tone of two frequencies, $f_{\text{diff}} = f_2 - f_1$, can be perceived; for example, frequencies $f_2 = 1.1$ kHz and $f_2 = 1.0$ kHz produce the perception of a faint, low-frequency tone at 100 Hz.

An interesting kind of combination tone is the *cubic difference tone*, $f_{\text{cubic}} = 2f_1 - f_2$. For example, frequencies $f_1 = 1.0$ kHz and $f_2 = 1.1$ kHz create $f_{\text{cubic}} = (2 \cdot 1.0 - 1.1)$ kHz = 0.9 kHz. Cubic difference tones are already generated at low levels of excitation, which is in contrast to the general behaviour of non-linear distortion. Therefore, studying non-linear cochlear mechanics with the cubic difference tone phenomenon has been of special interest.

7.5 Otoacoustic Emissions

When a sound is presented to the ear, an echo can be recorded. The echo can be delayed so much (> 10 ms) that it cannot be explained by passive mechanics of the ear but neural systems must be involved. The resulting motion causes a faint sound, which is propagated back out of the ear and can be registered after stopping the acoustic stimulus.

The sounds are often weak, but they can be captured using sensitive microphones. The measurement of these emissions provides a means to research the functioning of the cochlea. These responses are called *otoacoustic emissions* or cochlear echoes. The echo can be related to the input signal frequency, it can be a combination tone, or a spontaneous generation of cochlear vibration. This is a behaviour characteristic of a healthy ear (Geisler, 1998). The subject is discussed in slightly more detail in Section 19.4.6.

In many hearing impairments, the active role of the cochlea is damaged and the echo cannot be detected. Such measurements are commonly used in the diagnostics of hearing. The first test of the hearing of newborn babies in some countries is done by measuring the level of otoacoustic emissions. A clear response is taken as evidence of a functional inner ear.

The ear also emits sound during silence. The level of emissions is typically lower than 5 dB SPL. Although such spontaneous emissions have no great importance in the research on hearing, some interesting details of hearing have been found to be connected to them. For example, subjects with strong sinusoidal acoustic emissions have higher sensitivity at frequencies close to emissions (Heise *et al.*, 2009).

7.6 Auditory Nerve

The basic building block of the nervous system is the neuron. Neurons are cells specialized in receiving, processing, and storing information via electrical and chemical means. A typical neuron consists of *dendrites*, the *soma*, and an *axon* (see Figure 7.15). The neuron receives input from other cells through its dendrites. This input converges in the soma, from which the output is sent to other neurons through the axon. The meeting point of the axon and the
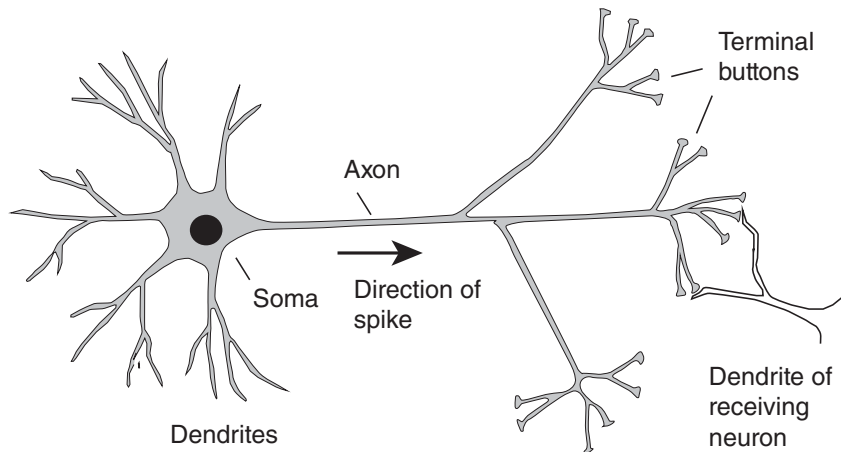
Figure 7.15 A neuron with a structure found typically in the brain. A multitude of dendrites conduct the action potential originating from other neurons to the soma. The soma processes the input and fires a spike through its axon, which then branches to reach the dendrites of receiving neurons.

dendrite of another neuron has a *synapse*, a tiny gap between the two neurons across which they communicate using chemicals called *neurotransmitters*. The axon releases these and they are received by the dendrite. The neurotransmitters modulate the membrane potential of the receiving neuron. These gradual fluctuations in the membrane potential progress to the soma. If the membrane potential reaches a threshold value, it gives rise to an *action potential*, which is an all-or-nothing cascade of events that includes a sudden large change in membrane potential followed by its return to the resting potential. This action potential then travels along the axon and triggers the release of neurotransmitters to the next neurons.

In the cochlea, the activity of the hair cells modulates the membrane potential of spiral ganglion neurons. The axons of the spiral ganglion neurons form the *auditory nerve* along which information travels from the cochlea to the brain in the form of action potentials. The vast majority of auditory nerve fibres transmit information from the inner hair cells, about 20 from each inner hair cell (see Figure 7.8). The rest of the fibres in the auditory nerve connect to the outer hair cells, and their role is not well understood (Plack, 2013).

Action potentials, often also called *spikes*, can be detected with the methods of neurophysiology. The rate at which the action potentials occur is often called the firing rate. A tiny recording electrode can be surgically placed in the proximity of the nerve. Neural responses in the auditory nerve fibres can be registered for different stimulus signals sent to the ear, and thus the whole chain of peripheral hearing can be investigated experimentally. Since the system is not linear and time-invariant (LTI), no single transfer function can represent it, but many points of view must be studied and characterized with different input–output relationships.

7.6.1 Information Transmission using the Firing Rate

The auditory nerve fibres show spontaneous firing even when no sound enters the ear. When the sound level is increased, the firing rate grows monotonically but starts to saturate above a specific level, as shown in Figure 7.16. The dynamic range is relatively narrow, about 20–40 dB for a single neuron (Moore, 2012). Auditory nerve fibres with a high spontaneous firing rate (about 20/s or more) are more sensitive but saturate first. The majority of neurons are of this type. Fewer medium spontaneous firing rate neurons also exist, and they are less sensitive but
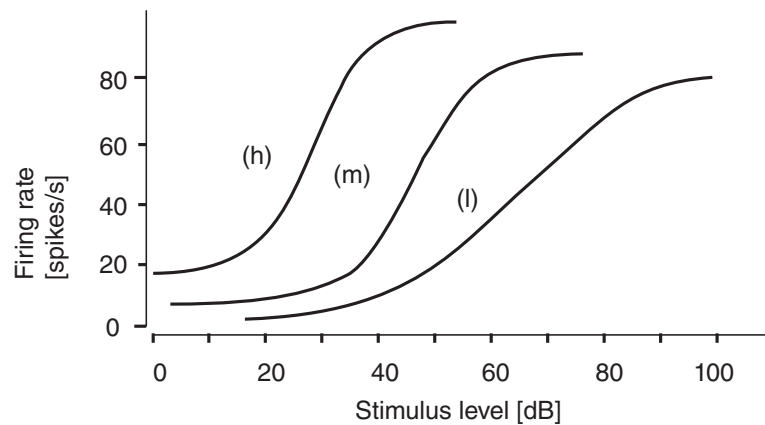
Figure 7.16 Level–rate functions for three neurons with different spontaneous firing rates: high (h), medium (m), and low (l). The level of stimulus is in dB SPL and the rate is in spikes per second. Adapted from Pickles (1988).
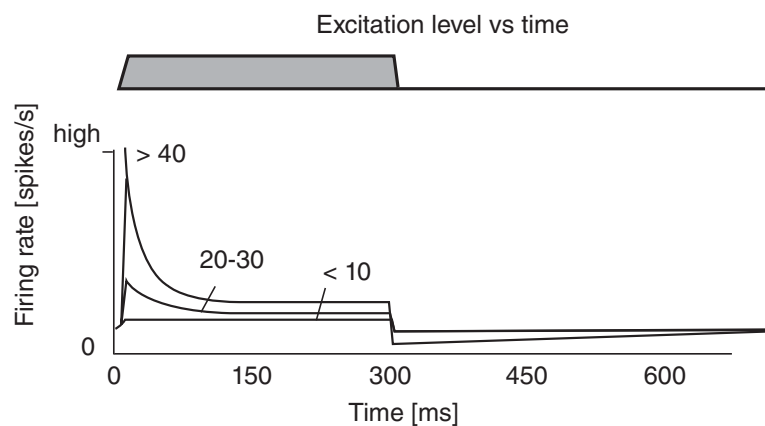

Figure 7.17 A schematic representation of a post-stimulus time histogram of the firing rate for a 300-ms long excitation at different levels. The curves exhibit overshoot after the onset and undershoot after the end of the excitation. The upper curve shows the time period of excitation. Lower curves show responses for different excitation levels (given as a parameter in dB).

work for higher levels of sound. The rare low-spontaneous-rate neurons remain unsaturated at high levels.

The output of the neuron to the auditory nerve is practically constant for signal levels above 50–60 dB, and the question then remains, how does our hearing analyse the loudness of sound at higher levels? A possible answer is that the cross-frequency spreading of the activation caused by loud sounds shown in Figure 7.12 is taken into account. With such an assumption, the auditory system analyses the auditory input across frequency and combines the excitation in several adjacent bands to form the loudness perception.

The static level–rate functions of Figure 7.16 exhibit the limited range and non-linearity of neural encoding, but this is just one viewpoint. Figure 7.17 shows a more dynamic view, called the *post-stimulus time histogram* (PST histogram), where the firing rate is plotted as a function of time for a constant excitation lasting 300 ms for different excitation levels. There is a significant *overshoot* after the onset of the excitation, but not for very low levels. This overshoot greatly exceeds a static saturation level, and it emphasizes attacks and onsets that
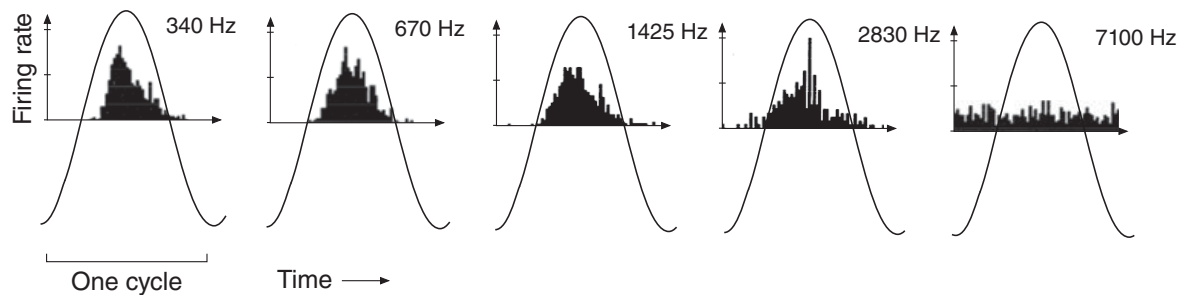
Figure 7.18 Post-stimulus time (PST) histograms for sinusoidal stimulation at different frequencies measured from a cat (data from Joris *et al.*, 1994). The neural firing in the auditory nerves at their best frequencies has been recorded for a continuous sinusoidal stimulus, and the results are summarized as PST histograms, which can be interpreted as the average activity of a neuron depending on the phase of the stimulus. The best frequencies of the nerves are specified. The sinusoidal stimuli are overlaid so that the phases match with the histogram. Reproduced with permission from The Acoustical Society of America.

are important from a perceptual point of view. When the excitation ends, there is an *undershoot* phenomenon before the spontaneous firing rate is recovered (Pickles, 1988).

7.6.2   *Phase Locking*

In some cases the neurons synchronize the firing with a certain phase of a periodic signal, a phenomenon called *phase locking*. This was observed in an experiment where the neurophysiological response from the auditory nerve of a cat to a periodic excitation was recorded, and the data were presented as a stimulus level versus synchronous firing rate plot, as shown in Figure 7.18 for different frequencies of a sinusoidal signal. The neural activity waveform is approximately a half-wave rectified (limited to positive values) form of the excitation waveform. In humans the synchronization clearly occurs at frequencies below 1500 Hz, degrades at frequencies above it, and is weak above 4 kHz. For such continuous signals, it is mainly the place and average rate that are mediated above this frequency, not the temporal details.

Most of neural processing is probabilistic and inaccurate when considering a single neuron, but neural signal processing is based on a large number of neurons working together, which can make the output of such neural networks very precise and robust.

The waveform of the signal also affects the synchronization, and phase locking does not necessarily occur at the same phase of each frequency component. With certain periodic sounds, the neurons in the cochlea synchronize to a specific phase of the fundamental frequency of the sound instead of the phase of the sinusoid(s) near the CF of the neuron. It has also been found that neurons with high characteristic frequency can phase lock to sinusoids with lower frequency.

So far we have not discussed the frequency selectivity of the neural responses. Since the basilar membrane (together with the outer hair cells) is able to do a selective-frequency analysis (frequency-to-place coding), it is reasonable to assume that frequency selectivity also exists in the auditory nerve coding. Figure 7.19 illustrates what happens when vowels are used as stimulus signals with varying levels. At low excitation levels, the neural firing in each characteristic frequency band shows a response that is approximately the vowel spectrum. The major formant ranges are easily seen (marked with arrows). At higher excitation levels, however, the
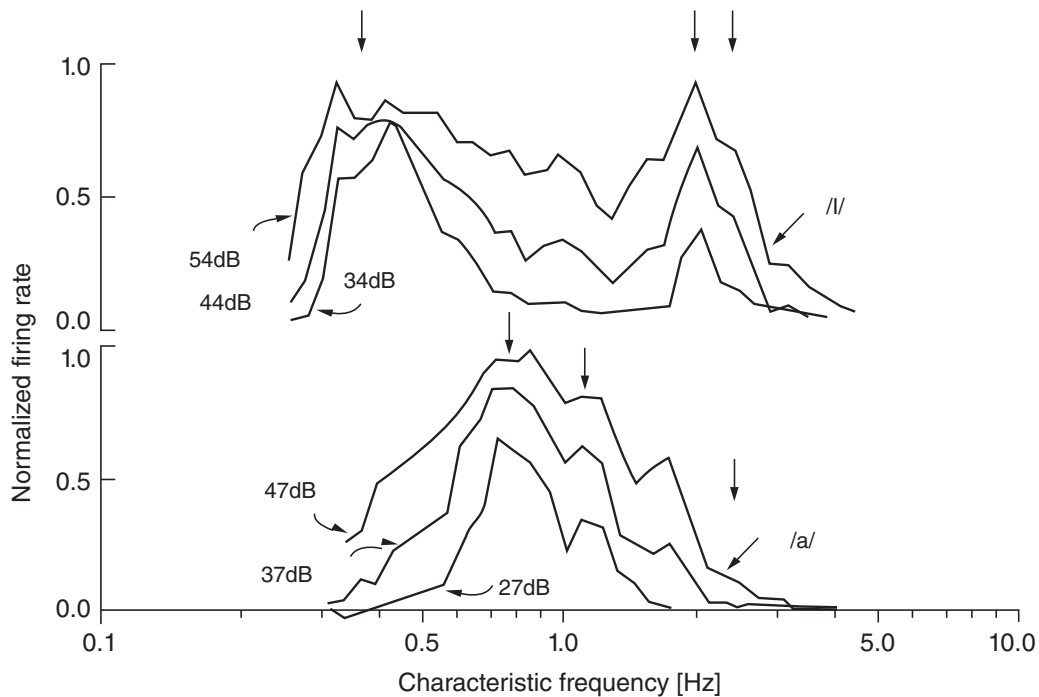
Figure 7.19   The averaged firing rate pattern produced for two vowels at different intensity levels measured from the auditory nerve of a cat. The higher stimulus level shows increasing saturation. Arrows indicate the main formant areas. Adapted from Sachs and Young (1979) and reprinted with permission from The Acoustical Society of America.

neural saturation effect distorts the spectral information, as seen in the frequency pattern of the firing rate (Sachs and Young, 1979).

   From Figures 7.16 and 7.19 one may conclude that the frequency resolution of hearing might be lost at levels above 60 dB due to neural saturation. This is not true, as we know from our everyday experience that speech with levels of even 100 dB is still understandable. An evident source of information is the temporal pattern of neural activation, and indeed Young and Sachs (1979) show that a combination of spectral and temporal neural patterns contains the information of vocal formants even at high frequencies.

7.7   Auditory Nervous System

The auditory nerve carries information from the cochlea to the brain. The number of neurons in the human brain is humongous, of the order of 100 billion (100 000 000 000) (Goldstein, 2013). The great number of neurons, and the even greater number of connections between them, enables the complex functions the brain conducts all the time. These neurons and their connections are packed very tightly in the brain in an extremely complex organization. It is thus a real challenge for science to figure out what the individual tasks of the neurons are and what kind of overall function the individual neurons conduct together.

7.7.1   Structure of the Auditory Pathway

Figure 7.20 is a highly schematic organization diagram of the central auditory system showing only the main units of neural processing and only some of the most obvious ascending connections. The neural system is divided into the left and right *hemispheres*, which are roughly
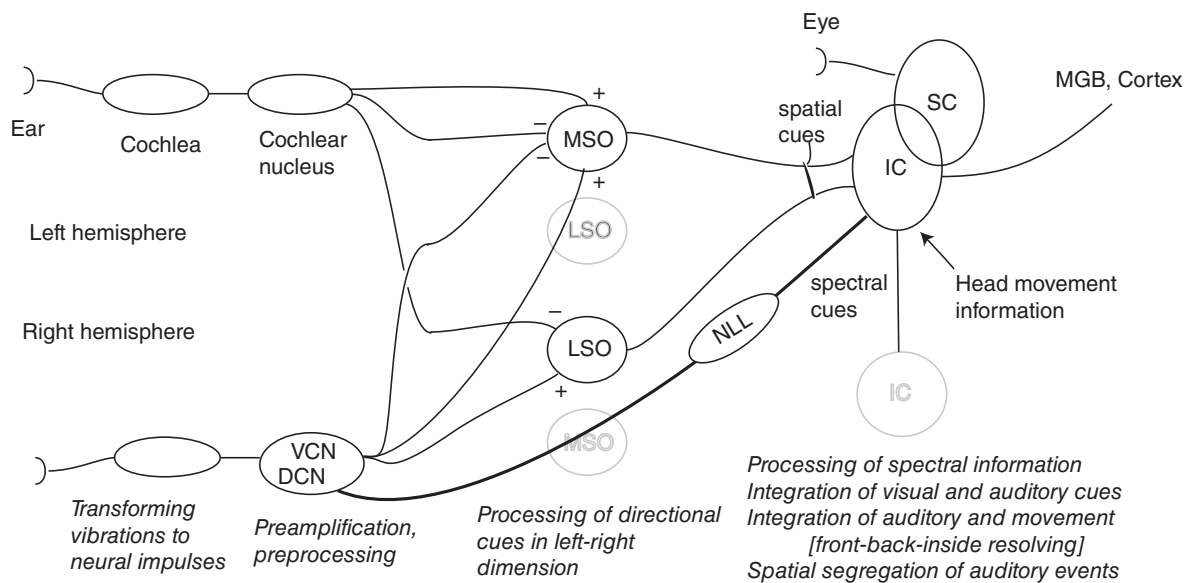
Figure 7.20   The most important nuclei and the auditory cortex as an organization of the central auditory nervous system. A simplified view of the most prominent ascending paths is shown.

symmetric with each other. All ascending fibres of the auditory nerve lead to the cochlear nucleus (CN), which is located in the brainstem. A plethora of neurons of different types exist, all performing different tasks. In engineering terms, the nucleus seems to act at least as a preamplifier, a signal router, and an initial periodicity and initial spectral analyser. The auditory pathway from the CN onwards to the inferior colliculus (IC) seems to be divided into two: one route leads to the IC directly or through the lateral lemniscus (LL) and the other through the superior olivary complex (SOC).

The neurons in the dorsal CN and in the postero-ventral CN give rise to the first pathway. The neurons have different forms, and they seem to perform a first-order analysis of the spectral content of sound, with some of them also phase locking to the repetitive structures in the time-domain signals, possibly affecting pitch perception (Schnupp *et al.*, 2011). The axons of these neurons lead directly, or through the LL, to the IC.

The antero-ventral CN, in turn, gives rise to the second route. It projects its output to the superior olivary complex (SOC). This connection is known for its accurate phase locking, exceeding the temporal accuracy found in the auditory nerve (Joris *et al.*, 1994). The enhancement is explained by the fact that CN neurons are innervated by multiple auditory nerves. The SOC is divided into the medial superior olive (MSO) and the lateral superior olive (LSO), both of which receive input from both hemispheres. Both the MSO and the LSO are known to be sensitive to interaural differences of ear canal signals (Grothe, 2003; Tollin, 2003). This processing thus enables the perception of left/right direction. The SOC projects to the IC directly or through the LL.

Numerous studies have measured the responses of IC neurons, but no single function has emerged for them. This may be related to the versatility of the IC: it is likely to perform a great variety of tasks. The IC transmits the auditory signals from the CN, LL, and SOC through the thalamus to the auditory cortex and also to the superior colliculus (SC) located next to it. The ICs in the two hemispheres are connected as well. A clear tonotopic organization can be found in the CN, SOC, LL, and in the IC.

The SC is also interesting in the context of hearing. It receives both visual and auditory signals, and it has been found to be one of the organs responsible for cross-modal interaction (Møller, 2006). One of its tasks is orienting the head and gaze towards the stimuli. Neurons in the SC form a topographical map of spatial locations; the neurons are organized according to the angle of incident of sound and/or light that they are most sensitive to. The SC includes a map of the auditory space that is aligned with the visual map in such a manner that the neurons responsive to auditory or visual stimuli from a certain direction are found close to one another. Some SC neurons also respond to multimodal stimulation originating from the same spatial location (Meredith and Stein, 1986). Interestingly, the topographical organization of the auditory space has been found only in the SC of all brain nuclei in the mammalian auditory system.

As the neuronal processes in the cochlea are only partly understood, it is quite understandable that the knowledge of neural processes beyond the cochlea is even more sparse. However, for acoustic engineering applications, understanding the functioning of the cochlea is usually enough, although certain spatial audio reproduction methods rely on knowledge of the function of the nuclei in the SOC.

From the IC, the auditory information is transmitted through the medial geniculate body of the thalamus to the auditory cortex. The tonotopical organization is maintained throughout the auditory pathway and is also seen in the core areas of the auditory cortex. The auditory cortex comprises the core areas with the tonotopical maps and surrounding areas in which neurons have increasingly complex response properties. Processing in the auditory cortex is also thought to be organized into two streams. The 'what' stream is thought to perform the analysis of the sound spectrum ('what is the source?'), whereas the 'where' stream is thought to analyse the spatial information of different sound events in the auditory scene ('where is the source?') (Rauschecker and Tian, 2000). A similar processing scheme is found in the neural analysis of the visual pathway.

There is no clear endpoint to the auditory pathway. The auditory cortex is interconnected to other areas of the cortex. Activity related to sound processing is also seen outside the auditory cortex, especially during listening tasks that involve attention, memory, and multimodal processing.

7.7.2   Studying Brain Function

The action potentials of single neurons can be recorded using tiny electrodes placed near the neuron or even inside it so that fluctuations in the membrane potential can also be monitored. These studies involve some invasive procedures and are therefore normally conducted only on animals. In such neurophysiological studies, sounds are presented to the animal, and the activity of single neurons or groups of neurons is then recorded. By carefully controlling the stimulus, and with a high number of repetitions, the representation of the auditory signals and analysis performed by the neurons can be figured out.

There is a variety of non-invasive brain-imaging techniques for studying human brain function, where the activity of the brain is recorded through the skull. These methods record the activity of entire brain areas, not single neurons. They are also best suited for recording the activity of the cortex. In *electroencephalography* (EEG), electrodes are placed on the scalp, and the electric potential in relation to a reference voltage measured from the subject is recorded with time. The neural activity during and after the stimulus causes small changes in the recorded voltages. The benefit of EEG is that it responds quickly to changes

in brain activity. However, localizing the sources of activity and differentiating between the contributions of brain areas is difficult using EEG. Another, somewhat similar method is *magnetoencephalography* (MEG), where the magnetic fields caused by electrical currents in neurons are recorded with very sensitive magnetometers. It has better resolution in localizing the activity within the brain than EEG has, because it suffers less from distortions caused by the scalp and skull.

The anatomical structure of the living human brain can be mapped using magnetic resonance imaging (MRI). Functional MRI (fMRI) is also able to map the functional properties of the brain. fMRI records local changes in the metabolism of the brain that are related to neuronal activity. However, the temporal resolution of fMRI is in the order of only hundreds of milliseconds even at its best. fMRI scanners are also very noisy and therefore their use for studying the auditory system sets a special challenge. fMRI has, however, good spatial resolution and is well suited for identifying which brain areas are active.

7.8 Motivation for Building Computational Models of Hearing

As discussed above, studying the functioning of hearing is extremely difficult for many reasons, and all ethical direct or indirect approaches giving any clues on the functioning are valuable. An additional tool for this purpose is to build functional models of the auditory pathway called *auditory models*. The success or failure of the models then helps researchers understand the principles and details of the processing in the brain.

Human auditory perception has been modelled at many different levels. The principle is to build computational models of acoustic, physiological, or neural mechanisms that estimate the relation of perceived auditory attributes resulting from the presentation of stimuli with certain acoustic attributes. Ideally, the ear canal signals are fed into the system, and the auditory attributes that subjects perceive, on average, are estimated without bias or noise.

There are different motivations for building such models. First of all, being able to model the functionalities of different parts of the system provides basic knowledge of the auditory system. Often, more than one mechanism affects the perceived characteristics of an auditory object, making the analysis of the results of psychoacoustic tests complicated. The complications can be mitigated when the effects of the mechanisms can be isolated in the computational domain, making the interpretation of the results easier. Second, a very important application for such models is the evaluation of audio quality. Modern acoustic systems comprise various transfers in the acoustic, signal processing, and telecommunication domains. The measurement of quality of reproduced audio cannot be performed with simple linear measures, which calls for the usage of models of human hearing for such tasks. Ideally, the fidelity of audio systems could be estimated by computing the auditory cues reaching the conscious level. The third application of auditory models is in the diagnostics of hearing-impaired patients. The symptoms of the patient would be characterized with specific psychoacoustic tests, and the attributes of the model of human hearing were adjusted until the symptoms were explained. Optimally, this would track down the reason for the symptoms.

In principle, modelling could be performed by first modelling one neuron, followed by modelling the next neuron, until the entire auditory pathway was modelled. Unfortunately, this approach is doomed for two reasons: the current knowledge of the complex electrochemical functioning of the neurons and of the neuroanatomy of the pathway is far from the stage where all the details can be modelled. In addition, the enormous number of neurons devoted to hearing

makes it computationally out of reach in the near future. The present status of auditory models is discussed in more detail in Chapter 13.

Summary

This chapter served the purpose of providing a brief introduction to the structure and function (physiology) of hearing: how sound waves from the surrounding air enter the auditory system through the external, middle, and inner ear. In the inner ear, the sound signals are divided into frequency channels and coded as neural spikes. The research methods used and the structure of the auditory pathway were also briefly introduced. Computational modelling of the mechanisms can then be used to supplement and test the knowledge obtained from neurophysiology.

Further Reading

Neural systems and their research is introduced in (Gelfand, 2004; Kalat, 2011). The methodology of investigating the physiology of the inner ear is described, for example, in Pickles (1988) and Yost (1994). There exist a lot of sources on the neuroanatomy and the neurophysiology of the auditory pathway. The reader might consider Popper and Fay (1992) or Webster *et al.* (1992) as an interesting starting point.

References

Aibara, R., Welsh, J.T., Puria, S., and Goode, R.L. (2001) Human middle-ear sound transfer function and cochlear input impedance. *Hearing Res.*, **152**(1), 100–109.

Arthur, R., Pfeiffer, R., and Suga, N. (1971) Properties of two-tone inhibition in primary auditory neurones. *J. Physiology*, **212**(3), 593–609.

Delgutte, B. (1990) Two-tone rate suppression in auditory-nerve fibers: Dependence on suppressor frequency and level. *Hearing Res.*, **49**(1), 225–246.

Geisler, C.D. (1998) *From Sound To Synapse: Physiology of the Mammalian Ear*. Oxford University Press.

Gelfand, S.A. (2004) *Hearing: An introduction to psychological and physiological acoustics*. Marcel Dekker.

Goldstein, E.B. (2013) *Sensation and Perception*, 9th edn. Cengage Learning.

Grothe, B. (2003) New roles for synaptic inhibition in sound localization. *Nature Reviews Neuroscience*, **4**(7), 540–550.

Heise, S.J., Mauermann, M., and Verhey, J.L. (2009) Investigating possible mechanisms behind the effect of threshold fine structure on amplitude modulation perception. *J. Acoust. Soc. Am.*, **126**(5), 2490–2500.

Joris, P.X., Carney, L.H., Smith, P.H., and Yin, T. (1994) Enhancement of neural synchronization in the anteroventral cochlear nucleus. I. responses to tones at the characteristic frequency. *J. Neurophys.*, **71**(3), 1022–1036.

Kalat, J.W. (2011) *Biological Psychology*. Cengage Learning.

Meredith, M.A. and Stein, B.E. (1986) Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J. Neurophysiol.*, **56**(3), 640–662.

Møller, A.R. (2006) *Hearing: Anatomy, Physiology, and Disorders of the Auditory System*. Elsevier.

Moore, B.C.J. (2012) *An Introduction to the Psychology of Hearing*, 6th edn. Brill.

Pickles, J.O. (1988) *An Introduction to the Physiology of Hearing*. Academic Press.

Plack, C.J. (2013) *The Sense of Hearing*. Psychology Press.

Popper, A.N. and Fay, R.R. (1992) *The Mammalian Auditory Pathway: Neurophysiology*. Springer.

Rauschecker, J.P. and Tian, B. (2000) Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc. Natl Acad. Sci.*, **97**(22), 11800–11806.

Ruggero, M.A., Rich, N.C., Recio, A., Narayan, S.S., and Robles, L. (1997) Basilar-membrane responses to tones at the base of the chinchilla cochlea. *J. Acoust. Soc. Am.*, **101**(4), 2151–2163.

Sachs, M.B. and Young, E.D. (1979) Encoding of steady-state vowels in the auditory nerve: Representation in terms of discharge rate. *J. Acoust. Soc. Am.*, **66**, 470–479.

Schnupp, J., Nelken, I., and King, A. (2011) *Auditory Neuroscience: Making Sense of Sound*. MIT Press.

Tollin, D.J. (2003) The lateral superior olive: A functional role in sound source localization. *Neuroscientist*, **9**(2), 127–143.

Verhulst, S., Dau, T., and Shera, C.A. (2012) Nonlinear time-domain cochlear model for transient stimulation and human otoacoustic emission. *J. Acoust. Soc. Am.*, **132**(6), 3842–3848.

von Békésy, G. (1960) *Experiments in Hearing*. McGraw-Hill and Acoust. Soc. Am.

Webster, D.B., Popper, A.N., and Fay, R.R. (1992) *The Mammalian Auditory Pathway: Neuroanatomy* volume 1. Springer.

Yates, G.K. (1995) Cochlear structure and function. *Hearing*, pp. 41–74.

Yost, W. (ed.) (1994) *Funfamentals of Hearing – An Introduction*. Academic Press.

Young, E.D. and Sachs, M.B. (1979) Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. *J. Acoust. Soc. Am.*, **66**, 1381–1403.