Sensation, Perception, Action – An Evolutionary Perspective

Chapter 7

Hearing 1: Sound and Noise

‘Angels Playing Music; Ghent Altarpiece’ Jan Van Eyck, 1426-27, St Bavo, Ghent

Sound has a very special quality for humans, which often associated with heavenly creatures such as angels, and indeed it strongly affects our emotional balance. An attempt to demystify the nature of sound is based on understanding its physical nature, and the systematic composition of sound and noise from pure tones.

Overview

Why is hearing so important for humans? What were the evolutionary pressures, and what makes the auditory sensory system so powerful? The acoustic channel gives us access to vital information, it is essential as an alarm system, helps us orienting through our environment, and is the basis of communication by using simple signals and language. To understand the design of the auditory system, and to appreciate what we can hear and what we cannot hear, we have to begin with the physical nature of sound, and its transduction into neural signals as the starting point of the auditory stream into the auditory cortex. The extraordinary sensitivity and adaptability of the acoustic sense is made possible through an intricate mechanical design of the middle ear, and a sophisticated encoding of sound waves into electric signals in the inner ear, which includes neural sharpening of tuning through lateral inhibition and active processes to set the operating range. Auditory information is encoded in terms of frequency (tonotopically organised frequency channels) and amplitude (response rates of individual neurons that may be combined in a population code), corresponding to the perception of pitch and loudness. Hearing performance can be described accordingly by systematic measurements of auditory sensitivity as a function of frequency, which is represented by ‘equal-loudness contours’. The concept of receptive fields and tuning will be revisited in the acoustic domain, and frequency masking will be presented as an experimental paradigm to demonstrate the existence of frequency channels, and their relevance for the operation of the auditory system in natural environments. The exact measure of sound pressure (in dB) will finally put into relation with the perception of loudness in its ecological context of typical environments experienced by humans, and some practical consequences in the context of occupational and environmental health are discussed.

Why do you need ears, why should you listen?

The first question about auditory perception needs to be why hearing is important for us as human beings. Sounds signal events, and because we are interested in events that happen in our environment, hearing opens a powerful channel to collect a range of information from the outside world. The sound of an approaching train or braking car, a mouse rustling in the leaves, the impact of the snowball on a window, or raindrops tapping on a tin roof, are some random examples of how we use our ears to orient ourselves in a cluttered environment, and relate to other objects and creatures. Most importantly, important alarm signals are transmitted through the auditory channel, and many animals have developed highly specialised hearing mechanisms to pick up the sound of approaching predators, or trace animals of prey. But there is more to hearing, because we not only hear sound passively generated by other animate or inanimate objects, but we also generate sound to be heard by others. Just like the songs of birds, the courtship signals of insects, or alarm calls of higher mammals, humans communicate by producing acoustic signals. Acoustic communication has in humans developed into a highly sophisticated system, language, which requires a fine auditory system to receive and extract all the information contained in acoustic signals, sometimes hidden in very fine nuances. Because hearing is so essential for communication, and communication is crucial for social interaction, the auditory system has a particular importance for humans and in consequence is a highly developed part of the human sensory system. Some psychologists would even argue that as a sensory modality hearing is even more important than vision. This view is not only supported by the fact that a core aspect of the human condition is the ability to communicate through spoken language, but also by the observation that our emotional balance is strongly affected by sound – just think about how distressed you can become by loud noise, and how relaxing it is to listen to music.

So let us think for a moment what are the most interesting problems related to auditory perception, and what are the most challenging questions for the psychologist, who is studying sensory systems. Once the basic mechanisms of encoding and decoding of acoustic signals have been understood, there are a lot of exciting questions about more advanced stages of auditory perception, such as: How does the brain enable us to recognize spoken words and the voices of individual speakers? What are the mechanisms of separating independent signal sources in a mix of voices coming from a group of people or different melody lines coming from an symphony orchestra? What is the influence of experience and knowledge on auditory perception? What is the perceptual basis of harmony and why do some sounds are always perceived as nice and others as terribly unpleasant? Whereas some of the more challenging questions will be addressed in the next chapter, we first turn to basic mechanisms of encoding sound signals in the human sensory system, and will start this with discussing the physical nature of sound.

The nature of sound

Any sound source, be it an exploding balloon or the plucked string of a guitar, is emitting spherical pressure waves which are expanding from the sound source and travelling through the air in all directions, regular shells of air compression and rarification emerge from the origin of a sound (schematically illustrated in figure 7.1 a), similar to the concentrically expanding ripples on a pond when a stone was dropped into the still water surface (figure 7.1b). For the case of a pure tone the simple physical event of creating a regular pattern of increasing and decreasing air pressure – both in space and time –can be exactly and simply described by a travelling sinewave (Barlow and Mollon 1982). The sinewave is characterised by two parameters, its amplitude and frequency. The amplitude (see figure 7.1 c) is the distance between the maximum (peak) air pressure and the minimum (trough) air pressure; the amplitude determines how loud a tone is perceived. The frequency of a pure tone (see figure 7.1 c) is the number of pressure changes per unit of time, which is measured in cycles per second, or Hertz (Hz), or the inverse of the duration of the full cycle of a complete wave, the period from the highest pressure through the lowest pressure back to the highest pressure; the frequency determines whether a tone is perceived as high or low, i.e. its pitch (Plack 2005).

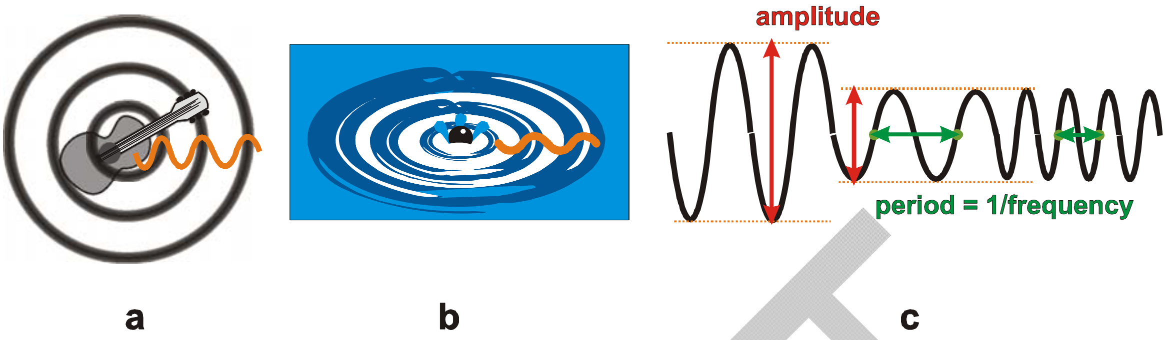

Figure 7.1: Sounds emitted from a musical instrument or other sound sources are spherical pressure waves in the air (a), similar to the radiating ripples on the water surface, when a pebble is tossed into a still pond (b). A pure tone is represented by a sinewave (c) travelling through space (air pressure as function of space or time), which is characterised by its amplitude and frequency (inverse of period).

This simple description of pure tones allows us to arrange them in a coherent system in a physical parameter space, which directly corresponds to their perceptual quality. The system is very well known because it is nothing else has the musical scale, which is learnt where everyone wants to master a musical instrument (Pierce 1992). Imagine, for instance, the arrangement of tones on the keyboard, which directly correspond to the notes on the score sheet for the piece of music that you want to play. Figure 7.2 illustrates how seven keys are arranged on the keyboard could generate a C-major diatonic scale. The rise in frequency of the musical tone generated by pressing the keys corresponds to the increasing pitch perceived by the listener. Obviously, this musical scale does not require a keyboard, but it can be produced on any musical instrument, as well as by the human voice. For singers, ‘solfeggio’ is often used to transcribe tones of the scale into syllables – just think of the famous song ‘Do-Re-Mi’ that Fraulein Maria uses to teach the musical scale to the von Trapp children in ‘The sound of music’. As you can see in figure 7.2, only very few of the tones are characterised by simple to remember integer frequencies, such as the A with 440 Hz. Also the relationships between the frequencies of consecutive tones are not linear (the differences tend to grow with increasing frequency). The reason for this is that the steps of the musical scale are not defined by constant differences in frequency, but by multiples, i.e. constant frequency ratios. An octave interval, the 7 full tone steps that lead to the next tone with the same name, for instance, corresponds to the doubling of the frequency. Similarly, a frequency ratio of 3/2 relates to a harmonic interval called a ‘fifth’ that spans 5 full tones, , and 4/3 to a ‘fourth’, etc. (Pierce 1992).

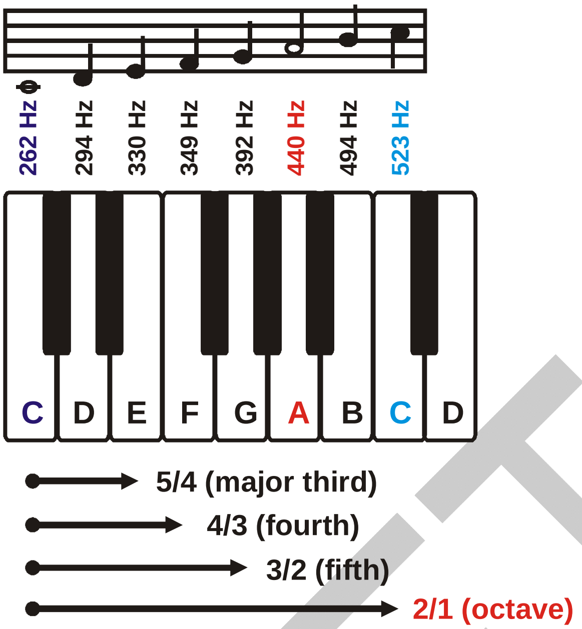

Figure 7.2: Musical tones as arranged on a keyboard (middle) correspond to notes on the musical score (top), and characteristic frequencies of the sound waves generated by pressing this key. In the diatonic scale shown here (C-major), tones are related to each other by characteristic frequency ratios, or ‘harmonic intervals’ (bottom). The tone ‘A’, corresponding to 440 Hz, is often used as absolute reference for tuning an instrument.

If the tone is unambiguously characterised by its frequency and amplitude, which correspond to perceived pitch and loudness, why is it that the same tone sounds so different and generated by different instruments? Indeed, there is more to a tone than just a simple sinewave (Evans 1982b). The crucial point is that musical instruments can not generate pure tones but produce complex waveforms that are composed of several sinewaves (illustrated in figure 7.3 for the superposition of two sinewaves). In order to generate a simple sinewave, or pure tone, you would need an electronic synthesizer as introduced to music production in the 1960s, and later developed into very powerful and now widely used tools to create electronic music that can sound truly natural (Russ 2004). Produced by a natural instrument, or indeed by the human voice, a tone with exactly the same pitch can sound so differently, because the sinewave that defines the pitch – the so-called ‘fundamental’, f_0 in figure 7.3 – is superimposed by a number of additional sinewaves with higher frequencies, which are integer multiples of the fundamental – the so-called ‘harmonics’; a single harmonic with $f_1 = 2 * f_0$ is shown in figure 7.3. The number, frequency ratios and relative amplitudes of the harmonics determines the exact waveform of the tone generated by a particular instrument, the so-called ‘frequency spectrum’, which is perceived as ‘timbre’, independent of the perceived fundamental frequency, its pitch. The outstanding capability of the human auditory system is the reason why to the trained ear it is immediately clear whether a tone A, for instance, was played on the piano, harpsichord, guitar, or xylophone. Now imagine the complexity of the auditory recognition task when more complex sounds are generated by adding further fundamental frequencies when playing a chord, or when a large-scale orchestra is working on a single bar of a symphony!

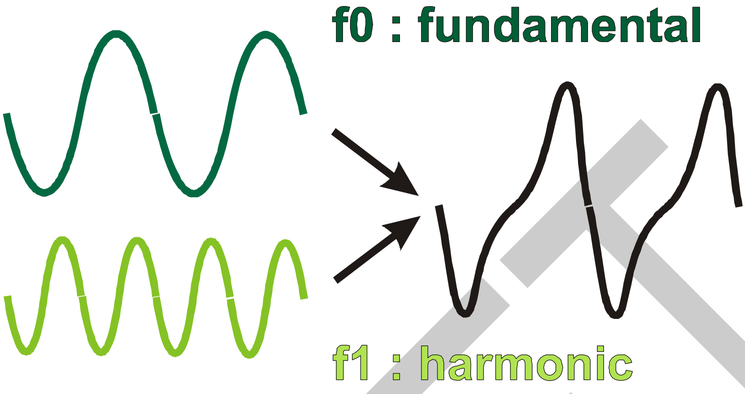

Figure 7.3: Natural tones generated by musical instruments are composed of the fundamental frequency f_1 (also called 'first harmonic'), perceived as pitch, and a set of harmonics (only the second harmonic $f_2 = 2 * f_1$ is shown here) which generate a complex waveform that is characteristic for the sound source and is perceived as timbre.

Finally, what happens when you superimpose random tones, or in other words produce a large number of pressure waves with independent properties at the same time? When a pure tone can be compared with throwing a pebble into a still pond, this is a situation resembles tossing a handful of pebbles into the pond, or raindrops falling on the water surface. At some point individual waveforms can no longer be distinguished and disappear into a irregular pattern of surface movement. The interference between the initial waveforms generating such random patterns in the sound space make it difficult to separate the individual contributions, or tones. The superposition of many tones with random amplitude and frequency sounds like irregular noise (Evans 1982b). This expression has a special meaning for information theory: the absence of any regularity in a signal means that it does not contain any information, which is called 'white noise', no matter whether it is an acoustic signal, a visual, or any other signal. What you hear and see when your television is detuned, i.e. as generating random auditory and visual signals, is white noise!

The ear: a highly sophisticated sensory organ for acoustic signals

The sensory organ to pick up auditory information is the ear, which works as a transducer, converting sound waves into neural signals. This biological system is characterised by an amazing functionality, which is not met by any artificial system engineered by humans. For instance, the ear is operating with outstanding sensitivity (Gulick et al 1989): the absolute threshold of hearing corresponds to sound levels that generate eardrum vibrations as small as 0.01 nm (less than the diameter of a hydrogen atom; the diameter of a human hair is a Million times larger!). The only visible part of the ear, the earlobe, or ‘pinna’ (sometimes also called ‘auricle’) arguably is the least relevant part of the ear, which mainly operates like a funnel to conduct the air pressure wave into the ear canal, leading to the eardrum, or ‘tympanic membrane’, that starts vibrating in the frequency of the air wave. Whereas the pinna and ear canal are referred to as ‘outer ear’, the section from eardrum to oval window is the ‘middle ear’, which is a fine mechanical structure known to many of us because it can fall victim to rather painful middle ear infections. The eardrum vibration is amplified in the middle ear, adapting it from the gas medium in the outer ear to the liquid medium in the inner ear, by means of three tiny bones known as the ossicles: the hammer, or ‘malleus’, the anvil, or ‘incus’, and the stirrup, or ‘stapes’ (Evans 1982a). This conversion process from low-power movement in air to high-power movement in fluid is called ‘impedance matching’.

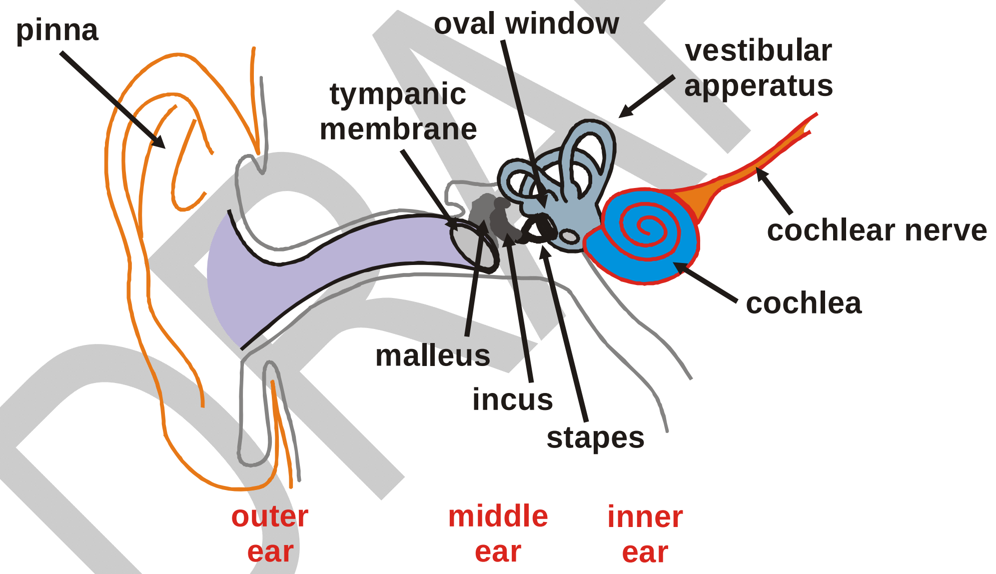

Figure 7.4: The ear is a sophisticatedly engineered sensory organ to pick up sound down to impressively low physical limits. The sound is conducted by the pinna into the ear canal leading to the eardrum (tympanic membrane), where it transformed by the three ossicles (malleus, incus, stapes) into movements of the fluid in the cochlea, behind the oval window. The mechanosensors in the cochlea transform movement into neural signals that are sent through the cochlear nerve (or ‘auditory’ nerve) into the brain.

The transformation of auditory information from mechanical oscillation into the neural code happens in the inner ear, which is the bony labyrinth containing a system of small cavities, or tubes that are filled with a viscous fluid. The two main functional parts of this system are (i) the vestibular apparatus, the organ of balance that consists of three semicircular canals and the vestibule, and (ii) the cochlea, which is the organ responsible for hearing, which contains the sensory surface that encodes sound. The mechanical stimulation is transmitted from the eardrum through ossicles onto the oval window, a membrane at the front end of the cochlea (Evans 1982a), where the oscillations are converted to pressure waves in the cochlea, which generates a travelling wave of the basilar membrane in the 'organ of Corti' (Plack 2005). The peculiar mechanical properties and their possible role for hearing were studied ingeniously by G. v. Békésy, who used microsurgery to observe the motion of tiny silver particles introduced into the cochlea (von Békésy 1961). He was awarded the Nobel prize for his discoveries.

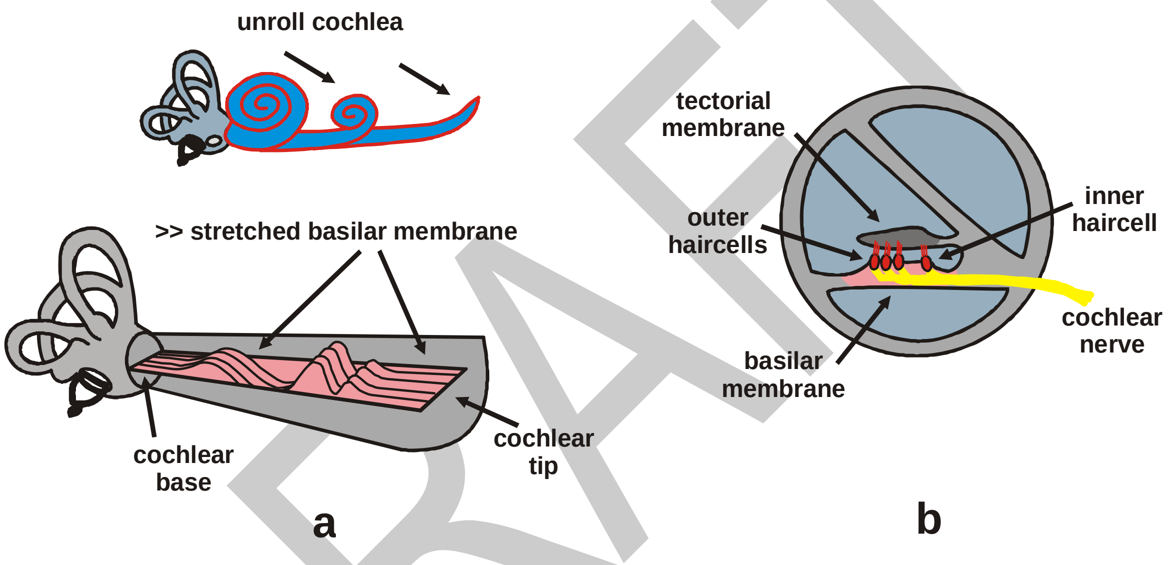

Figure 7.5: The part of the inner ear responsible for hearing is a helical structure, the cochlea, which can be seen in (a) as schematically un-coiled into an elongated tubular system. This segmented tube contains an extended 'basilar' membrane that is set into vibration (a travelling wave, similar to shaking a sheet, shown schematically in pink) by the sound transmitted by the ossicles through the oval window into the viscous fluid of the cochlea. Small haircells on the basilar membrane, shown in a cross-section across the cochlear tube in (b), are shifted relative to the tectorial membrane, and the bending of the cilia (hairs) leads to electrical activation. As a result, the travelling wave is transformed into neural signals that are sent through the cochlear nerve to the brain.

The cochlea is an extended canal system filled with viscous liquid that is coiled into a helical shape (see figure 7.5a), hence the name cochlea which is Latin for 'snail'. The essential part of this structure is the flexible basilar membrane separating two canal sections, which is covered with small hair cells that convert mechanical deformation, enhanced through a cover sheet, the ‘tectorial membrane’, into electric signals (see figure 7.5b) – the transduction process that is the crucial for all neural encoding in sensory systems. The full stretch of basilar membrane, together with its associated mechanosensor array of small haircells, constitutes the ‘organ of Corti’ inside the cochlea, which is responsible for hearing (Gopen et al 1997). One of its peculiar features of this design is that the vibrations of the basilar membrane are picked up by the hair cells and converted it into a space-specific representation of tones. The sound transmitted through the oval window into the cochlea fluid generates a travelling wave, an oscillation that moves along the length of the basilar membrane. Due to the mechanical properties of the membrane the wave is reaching a maximum amplitude at a location that is determined by the frequency of the stimulus. For small frequencies (low pitch tones) the largest oscillation is found at the apex of the cochlea, far away from the oval window, whereas large frequencies (high pitch tones) lead to the largest oscillation amplitude at the front end of the cochlea, close to the oval window (see figure 7.6a, top to bottom). Because the location of the strongest mechanical, and therefore strongest electrical signal, depends systematically on the frequency of the stimulus, the organ of Corti generates a map of decreasing pitch along the extension of the basilar membrane, from the oval window to the apex of the cochlea. In other words, a spatial code for stimulus frequency is generated (Evans 1982a). The nerve fibres that transmit the electrical signals from the haircells through the auditory or ‘cochlear’ nerve to the Cochlear Nucleus and other regions in the brainstem, and from there to the auditory cortex. These neurons carry information about sound intensity by the level of their activation, and sound frequency by means of their location being related at each stage to the region of origin on the basilar membrane. In the auditory cortex, neighbourhood relationships of haircells are conserved such that neurons are arranged systematically in order of increasing stimulus frequency (Formisano et al 2003), which leads to a ‘tonotopic map’ (from greek tonos = pitch, and topos = location) similar to retinotopic maps in the visual system.

Early in the study of hearing function it was noticed that the region of the basilar membrane that is oscillating when any pure tone is stimulating the ear is much more extended than the electrical activity of the haircells, which determines the range of frequencies that activate any particular nerve fibre in the cochlear nerve (von Békésy 1967b). This means that the frequency tuning of the mechanical system in the inner ear is broader than the frequency tuning in the neurons. So the information has been processed between the mechanical and the neural encoding, in a mechanism that some researcher refer to as ‘two-tone suppression’ (Moore 2003). The activity of a neuron in the cochlear nerve coding for a particular frequency is determined by the haircell activation from the centre of the mechanical oscillation, from which the haircell activation from the surround – the neighbouring regions further up and down the basilar membrane – is subtracted (see figure 7.6 b). Note that the fundamental mechanism in this sharpening of frequency tuning (von Békésy 1967a) shows properties that resemble the lateral inhibition that was discussed in the context of the visual system as basis of contrast enhancement and redundancy reduction through opponency (chapters 3 and 4).

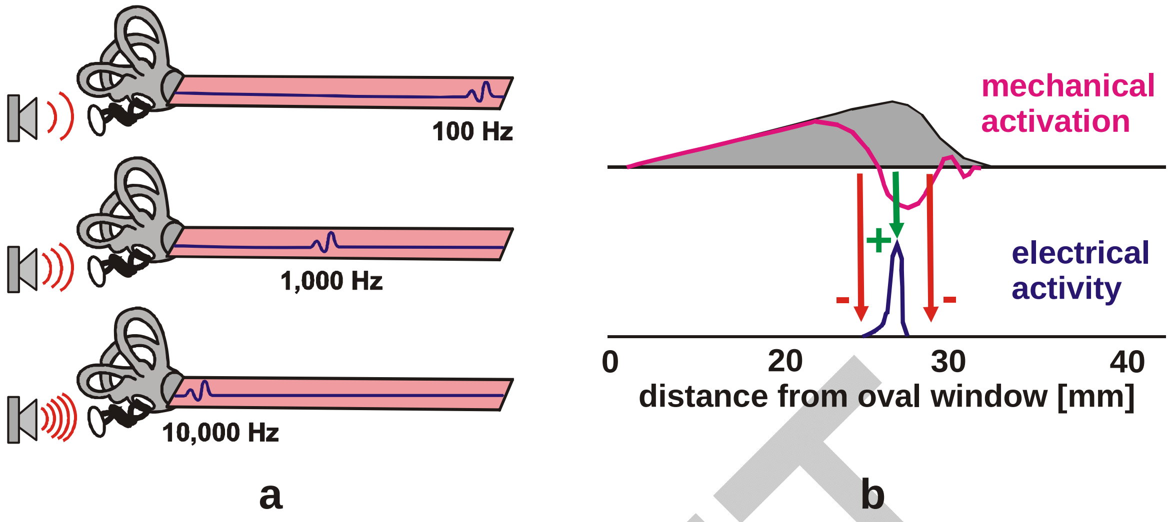

Figure 7.6: (a) Pure tones lead to oscillations in regions of the basilar membrane at locations that are determined by the tone’s frequency (from top to bottom: 100 Hz, 1,000 Hz, 10,000 Hz). (b) The region of mechanical activation (sketched as magenta deformation for one moment in time, and grey envelope as time average, shown schematically as function of distance on basilar membrane from the oval window) for a given tone is broader than the activation in the neurons (sketched schematically as blue profile) in the cochlear nerve transmitting these signals to the auditory cortex, which corresponds to the frequency tuning of the individual neuron. This sharpening of the tuning function is achieved through lateral inhibition, the activation from the surround (red arrows, “-“) is subtracted from the activation from the centre (green arrow, “+“).

The brief description of the cochlear function given here is no more than a coarse sketch, if not caricature, of what really happens in the ear when we are picking up sound – which we do for 24 hours a day, every day of our life. The actual electromechanical events in the ear, including some amazing mechanisms for amplification, selection and protection, are much more complex than could be described here (for a more detailed description, see Moore 2003) and they are still a matter of landmark research (Mellado Lagarde et al 2008). Furthermore, we neglected to discuss aspects of hearing that are not mediated by airborne sound signals but conducted through bone tissue (von Békésy 1949), which therefore is crucial for some hearing aids (Westerkull 2002). For our purpose it is sufficient to conclude that the encoding of acoustic signals requires several ‘engineering tasks’, which are accomplished by a biological system that has been driven by evolution to astonishing perfection. The outer ear serves as directional collector of pressure waves and conduit into the inner ear, which performs impedance matching, and protects against mechanical overload to the highly sensitive mechanical structures. The inner ear contains a sophisticated electromechanical transducer that includes the first steps of frequency analysis. On its way from the outside world into the cortex, auditory information is processed by the same fundamental mechanisms that we find in other sensory modalities, and have been discussing in some detail for the visual system. These mechanisms include (i) tuning, the preferential response of a sensor to a dedicated stimulus range (here tone frequency, or pitch); (ii) filtering mechanisms based on lateral inhibition, which sharpens frequency tuning to pitch in the peripheral auditory system; and (iii) sensory maps, the peripheral and cortical representation in which frequency information is ordered in its position in neural networks (tonotopy).

Pitch perception: frequency channels revealed by masking

So far we have looked at the anatomical and physiological aspects of hearing. But how can we study auditory perception, for instance the sensation associated with the frequency of tones, which is experienced as pitch? The most fundamental experimental method is called ‘masking’, which is a general technique used in many other areas of sensory or cognitive psychology. The phenomenon of auditory masking is schematically illustrated in figure 7.7 with an example from music, the phenomenon that you can only hear two separate instruments in an orchestra if they differ sufficiently in their frequency spectrum, and are adjusted in their volume such that the frequencies produced by one instruments (piccolo flute, in this example) are not completely eclipsed (or ‘masked’) by the frequencies produced by the other instrument (bassoon, in this case).

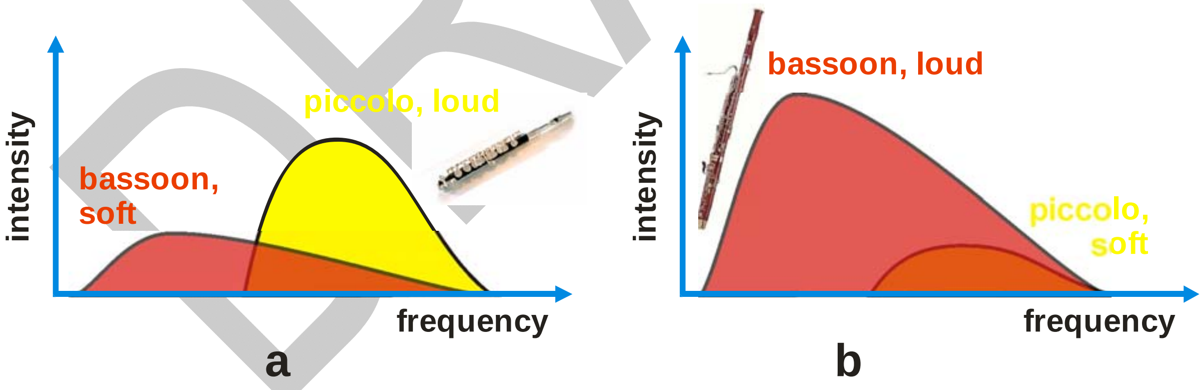

Figure 7.7: Masking is the perceptual disappearance of one stimulus (sound of the piccolo flute) in the sensory response generated by another, a different, stimulus (sound of the bassoon); the sound intensity of both instruments are plotted as a function of tone frequency. (a) You only hear the piccolo if the bassoon is played very softly. (B) If the bassoon is played at full strength, all its frequency components have a larger amplitude than those of the piccolo, and therefore completely conceal the piccolo’s frequencies, which no longer can be heard.

Based on this simple observation, a straight-forward and powerful auditory masking experiment can be designed. Participants, or listeners in this case, are asked to detect a target tone (for instance, a pure sine of 1000 Hz) in presence of a mask, which can be another tone with different frequency (Wegel and Lane 1924), or narrow-band noise, which is a small range of different frequencies with a particular centre frequency (Scharf 1971). The intensity of the masking sound is increased until the target tone can be no longer detected and decreased until the target tone is audible again, in order to determine the detection threshold for this target-mask combination. The mask intensity at detection threshold is inversely proportional to the sensitivity of the detection mechanism for this mask frequency, i.e. the signal intensity in the subset of neurons which is responsible for detecting the target tone. By repeating the same estimation procedure for a range of mask frequencies the sensitivity profile to stimulus frequency can be described, which is the ‘frequency tuning’ of the underlying neural mechanism. There are a number of variations of the target-mask conditions that can be used to measure different aspects of the encoding of tones in the auditory system, including the frequency interval that represents equal distances on the basilar membrane (cf. figure 7.6a), the ‘critical band’ (Greenwood 1961).

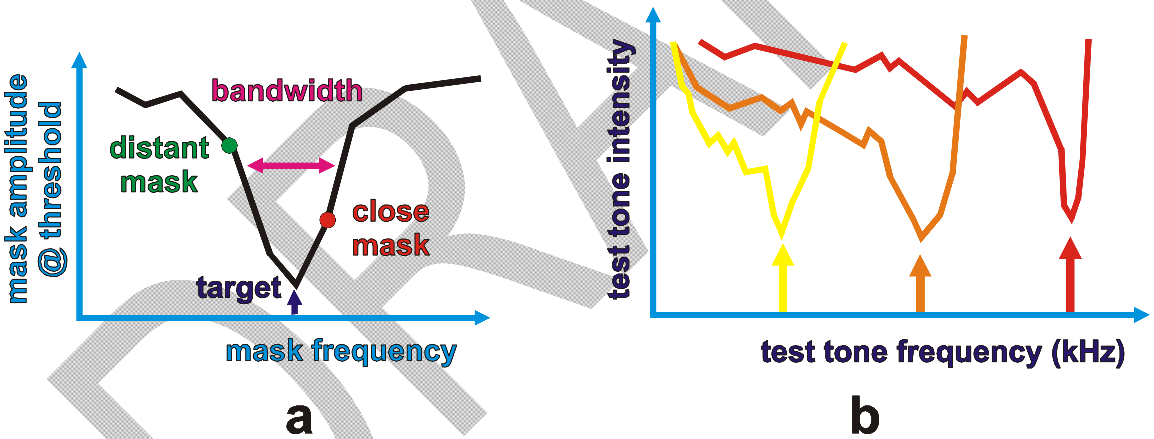

Figure 7.8: Frequency tuning. (a) Psychophysical measurements are used to determine the amplitudes of a masking sound at variable mask frequencies, which define the threshold for the detection of a simultaneously presented target tone. A mask close in frequency (red dot) requires a smaller amplitude than a distant mask (green dot), in order to suppress the detection of the target (blue arrow), leading to a U-shaped tuning curve (sketched schematically) that is responsible for frequency selectivity, with a characteristic location on the frequency axis and a particular bandwidth (magenta arrow). (b) In electrophysiological recordings (for instance, from the cochlear nerve) the threshold intensity of a test tone that is needed for eliciting a neural response can be used directly to describe frequency tuning of an auditory neuron

(schematically sketched here for three hypothetical neurons with different tuning frequencies, indicated by yellow, orange, red arrow).

By means of systematic variation of the mask frequency a set of detection thresholds is determined for a given target frequency to determine the psychophysical tuning curve for the neural filter mechanism responsible for this target frequency (Evans 1982a): when the mask amplitude at detection threshold is plotted as function of mask frequency (see figure 7.8a), low thresholds for masks close to the target frequency indicate high sensitivity of the filter tuned to these mask frequencies, and high thresholds for masks distant from the target frequency indicate low sensitivity. The minimum threshold (i.e. maximum sensitivity) of the U-shaped threshold curve points to the centre frequency of the filter in question (which is the target frequency in the masking experiment), and the frequency distance between the two points at half-height on the two slopes of the tuning curve is called the ‘bandwidth’ of the filter. The bandwidth is a measure for the sharpness of tuning – a small bandwidth means that only a narrow range of frequencies is processed by this filter (narrowband), whereas a large bandwidth indicates much less selectivity for stimulus frequency (broadband). Although this technical language initially may sound surprising in the context of biological systems and human perception, it reflects a practical and theoretical understanding of how pitch is encoded in the human auditory system: an array of neurons (i.e., set of filters) that respond preferentially to a set of different characteristic tones (i.e., are tuned to frequencies) represent pitch in the human auditory system, and are responsible for the detection and discrimination of different sounds (see figure 7.8b). We should also keep in mind that the terminology used to describe hearing is by no means a stranger to us – it is the same you will find in the manual of your CD-player, or stereo system, or digital TV, and the word ‘broadband’ arguably is the most popular word to use when it comes to your internet connection…

The physiological basis of pitch perception can be directly observed by means of electrophysiological recordings from individual fibres in the cochlear nerve (which is the analogue to the optic nerve in the visual system) – the axons of individual sensory neurons that transmit auditory information from the sensory organ to the auditory cortex. The sound intensity that is needed to generate activity in the nerve fibre is shown schematically in figure 7.8b as function of tone frequency for three units that are tuned to different frequencies (for real data, see Evans 1982a). It is immediately clear that each threshold U-shaped response profile can be directly compared to the frequency tuning measured psychophysically (figure 7.8a) with one target tone. A common way to represent the masking experiment results and the tuning properties of the neurons that mirror this behaviour is shown schematically in figure 7.9a as sensitivity profiles (the inverse of thresholds). In this form the curves are inverted from a U-shape to a Bell-shape, where the two more familiar curve parameters, midpoint and dispersion, may be helpful to interpret the meaning of the tuning parameters, centre frequency and bandwidth. Individual neurons tuned to different centre frequencies operate like an extended set of frequency-tuned filters that cover the full range of frequencies (sketched in figure 7.9b) as basis for frequency discrimination involved in auditory perception. Note that this similar strategy of frequency encoding is similar to the encoding that is called ‘Fourier analyser’ in the visual system (see chapter 3), and the same encoding specializing on noise reduction can be found in digital audio systems!

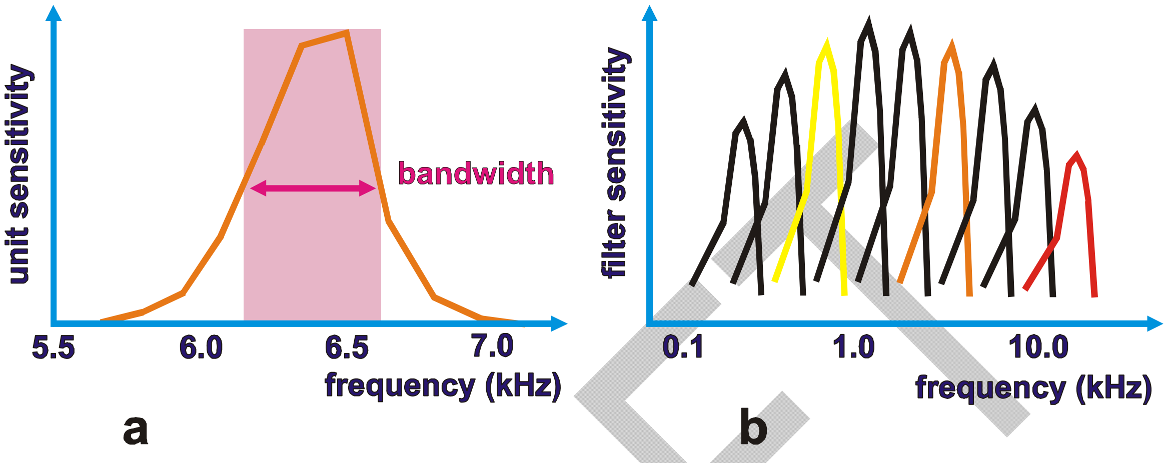

Figure 7.9: Frequency channels. (a) The tuning function of an individual filter, or channel, measured psychophysically or electro-physiologically, can be represented by a sensitivity profile as function of stimulus frequency with two critical parameters, centre frequency and bandwidth. (b) A set of several frequency channels characterises hearing performance, and can be used as basis to explain the detection of sound and frequency selectivity which is crucial for the discrimination of pitch.

The perception of loudness: measuring sensitivity to sound intensity

After having discussed the first aspect of hearing a tone, perceived pitch that is related to frequency, we now turn to the second parameter, the magnitude of auditory sensation which is determined by the pressure of airwaves and is related to perceived as loudness. Physically, sound intensity is defined by sound pressure level (Pierce 1992), SPL, and measured in decibels, dB, which is defined as logarithm of the ratio of the sample sound pressure, I, relative to the pressure of the normal hearing threshold at 1000 Hz (I_0 = 20 micropascals): SPL = 20 * log (I/I_0). This means that when sound pressure grows by a multiple of 10, the SPL value increases by 20 dB (see table 7.1). Whereas it initially seems simple to measure loudness, the subjectivity of such perceptual judgements is a substantial problem for such experiments. Nevertheless in several studies an astonishing level of consistency was found (Stevens 1956) when the intensity of tone is decreased or increased relative to a reference tone, and participants are asked to rate the relative magnitude (as multiples of the reference tone): subjective ratings of loudness change proportionally to SPL (see figure 7.10), suggesting a power law for this subjective estimate (Stevens 1957).

Stevens proposed to use ‘sones’ as units of perceived sound intensity, or loudness (Stevens 1936).

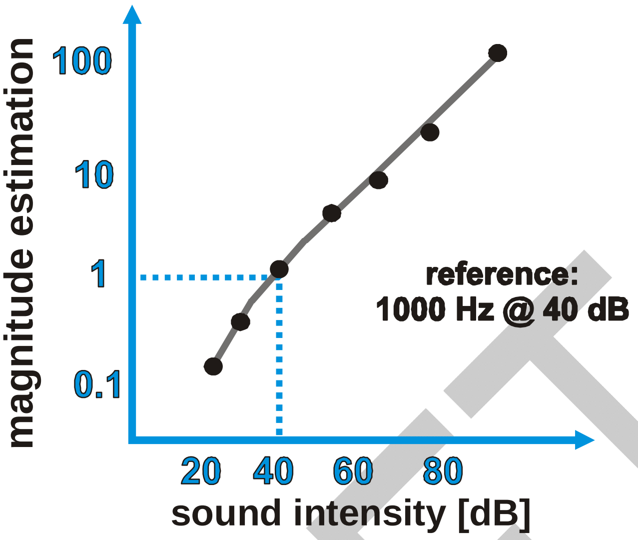

Figure 7.10: Perceived loudness. Participants are asked to estimate the magnitude of a test tone (for which SPL was varied) as multiples of a reference tone (in this example a 1000 Hz tone at 40 dB) – the relationship is approximately linear.

A method to measure perceived loudness quantitatively, which is not based on ratings and therefore is less dependent on the participant’s internal criteria and reliability, involves the comparison of two pure tones that are successively presented in a forced choice paradigm, by asking them to decide which one sounded louder. When one of the stimuli, the reference tone, is kept at constant frequency and SPL, and the SPL of the other stimulus, the comparison tone, is varied it can be determined when their loudness is perceived as equivalent (see sketch in figure 7.11a). It is common to use a 1000 Hz tone as reference, which is close to threshold at 0 dB SPL. When the SPL that are perceived as having the same loudness as the reference tone is determined for the full range of audible frequencies, contours of equal loudness can be generated (Berrien 1946). A set of such curves for several intensities of the reference tone leads to a family of equal-loudness contours (see figure 7.11b), including the absolute threshold contour which is the basis for the audibility function, or ‘audiogram’, which describes the limits of hearing performance for an individual as function of frequency, and therefore is widely used in the clinical practise to assess hearing impairments (Moore 2003).

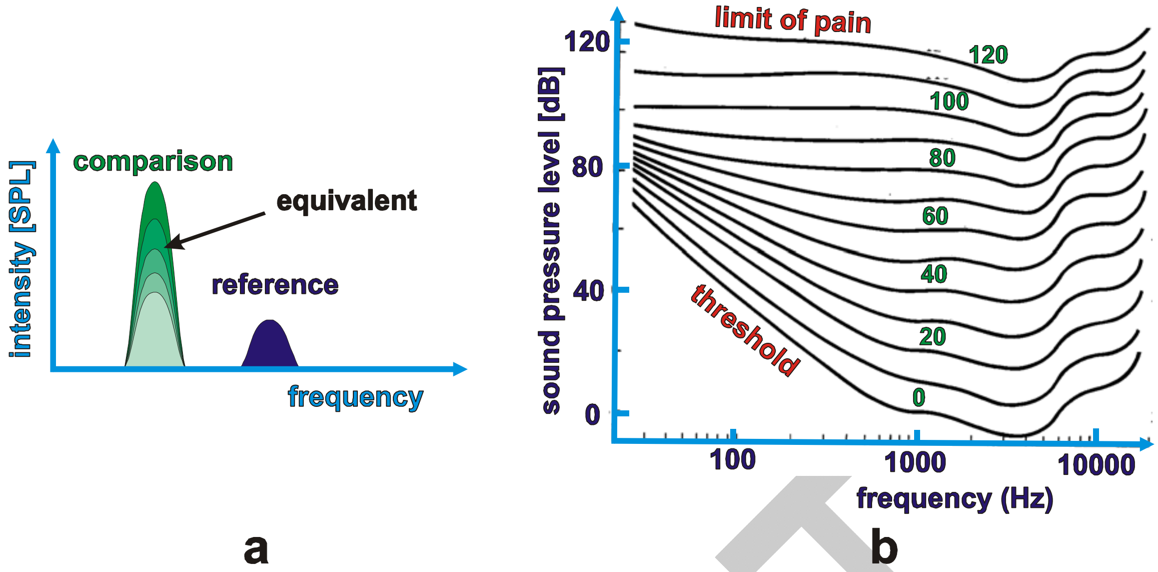

Figure 7.11: Measuring audibility and perceived loudness of tones at a range of frequencies. (a) the basic experimental design involves the evaluation of a reference tone of fixed frequency and SPL (blue Bell shape) and a comparison tone with variable SPL (green Bell shapes) to establish at which SPL they are perceived as being equally loud. (b) By testing a wide range of frequencies for the comparison tone in this matching paradigm, equal-loudness contours are generated for different intensities of the reference tone, between the threshold of hearing and high SPLs that approach the limit of pain.

Knowing that audible sound can vary in intensity by a factor of 10,000,000, or 140 dB, it is interesting to ask in which ecological contexts which sound intensities are experienced. As indicated by the set of examples listed in table 7.1, the sound pressure we are exposed to in natural environments are moderate (unless we are subjected to very dangerous situations, like an volcanic eruption), and high SPLs induce pain which would trigger escape reflexes. In man-made environments, however, we can be exposed to very high sound pressures that can damage the inner ear, leading to temporary (for instance, when enjoying techno music at close range, if you are lucky) or permanent hearing loss (explosions). Certain work places, such as airports, can present considerable occupational health hazards (May 2000), and protective gear is compulsory. A particular interesting, and highly debated, issue is how to measure and to assess the impact of chronic ambient noise on the population living in the neighbourhood of busy airports (Haines et al 2001) or on workers in noisy factories. Interestingly, the ear itself, being a biological system, has some protective and repair mechanisms to limit the impact of high pressure exposure (Robinson 2009), and intensive research is carried out to study the regeneration of cochlear haircells (Corwin and Oberholtzer 1997) that commonly were believed not to recover from damage in mammals.

	0 dB	threshold of hearing (at 1000 Hz)	
	10 dB	normal breathing	
10 x	20 dB	standing in a forest: leaves rustling in a breeze	
	30 dB	empty lecture theatre	
100 x	40 dB	College campus at night (without planes)	
	50 dB	quiet restaurant	
1,000 x	60 dB	two-person conversation	
	70 dB	standing at the roadside on Trafalgar Square	
10,000 x	80 dB	operating a vacuum cleaner	discomfort
	90 dB	getting close to a huge waterfall (Niagara)	**danger level**
100,000 x	100 dB	underground train passing the platform	**danger level**
1,000,000 x	120 dB	getting close to a propeller plane at takeoff	**hearing loss**
10,000,000 x	140 dB	Heathrow: standing near a jet plane at takeoff	**pain level**

Table 7.1 : The ecology of sound pressure. Examples of environments (third column) which expose a person to a range sound intensity (SPL), as measured in decibels (dB, second column) as logarithmic multiples of the hearing threshold at 1000 Hz; to help with this measure, the factors of sound pressure relative to threshold are given for 20dB changes in the first column. The last column indicates the risk of the sound levels in a given environment to damage the auditory system.

Take home messages

- the auditory 'channel' is an important source of crucial sensory information that opens another window to the world
- the physical nature of sound is based on waves of air pressure, which can be delivered as pure tones, composites, complex patterns, or noise

- the ear is a highly sensitive and intelligent device to pick up and convert sound pressure waves into electric signals; the essential transduction is based on the cochlear function in the inner ear
- frequency filtering, which originates from the electro-mechanical properties of the inner ear and is further enhanced by neural processing, is the basis of perceiving pitch; it can be studied electrophysiologically, or psychophysically in masking experiments
- sound intensity is the second characteristic property of sound, with high ecological significance, which is perceived as loudness
- hearing shares basic mechanisms with other sensory system, such as peripheral filtering, efficient neural encoding, and sensory maps in cortical representation

Discussion Questions

- Explain the physical nature of sound, with reference to the perception of pitch and loudness.
- How is tone frequency encoded in the auditory system?
- What is measured in the unit of ‘dB’, and what does it mean?
- Describe the impact of sound volume (loudness) with some examples from the real world.

References

Barlow H B, Mollon J D, 1982 *The senses* (Cambridge London New York: Cambridge University Press)

Berrien F K, 1946 "The effects of noise" *Psychological Bulletin* **43** 141-161

Corwin J T, Oberholtzer J C, 1997 "Fish n' Chicks: Model Recipes for Hair-Cell Regeneration?" *Neuron* **19** 951-954

Evans E F, 1982a "Functional anatomy of the auditory system", in *The Senses* Eds H B Barlow and J D Mollon (Cambridge: Cambridge) pp 255-306

Evans E F, 1982b "Functions of the auditory systems", in *The Senses* Eds H B Barlow and J D Mollon (Cambridge: Cambridge) pp 307-331

Formisano E, Kim D-S, Di Salle F, van de Moortele P-F, Ugurbil K, Goebel R, 2003 "Mirror-Symmetric Tonotopic Maps in Human Primary Auditory Cortex" *Neuron* **40** 859-869

Gopen Q, Rosowski J J, Merchant S N, 1997 "Anatomy of the normal human cochlear aqueduct with functional implications" *Hearing Research* **107** 9-22

Greenwood D D, 1961 "Auditory Masking and the Critical Band" *The Journal of the Acoustical Society of America* **33** 484-502

Gulick L G, Gescheider G A, Frisina R D, 1989 *Hearing: Physiological Acoustics, Neural Coding, and Psychoacoustics* (New York: Oxford University Press)

Haines M M, Stansfeld S A, Job R F S, Berglund B, Head J, 2001 "Chronic aircraft noise exposure, stress responses, mental health and cognitive performance in school children" *Psychological Medicine* **31** 265

May J J, 2000 "Occupational hearing loss" *American Journal of Industrial Medicine* **37** 112-120

Mellado Lagarde M M, Drexl M, Lukashkina V A, Lukashkin A N, Russell I J, 2008 "Outer hair cell somatic, not hair bundle, motility is the basis of the cochlear amplifier" *Nat Neurosci* **11** 746-748

Moore B C J, 2003 *An Introduction to the Psychology of Hearing* (San Diego: Academic Press)

Pierce J R, 1992 *The science of musical sound* (New York: Freeman & Co)

Plack C J, 2005 *The Sense Of Hearing* (Mahwah, NJ: Lawrence Erlbaum Associates)

Robinson R, 2009 "Feedback System Protects Inner Ear" *PLoS Biology* **7** e12

Russ M, 2004 *Sound Synthesis and Sampling* (Oxford: Focal Press)

Scharf B, 1971 "Fundamentals of auditory masking" *Audiology* **10** 30-40

Stevens S S, 1936 "A scale for the measurement of a psychological magnitude: loudness" *Psychological Review* **43** 405-416

Stevens S S, 1956 "The Direct Estimation of Sensory Magnitudes: Loudness" *The American Journal of Psychology* **69** 1-25

Stevens S S, 1957 "On the psychophysical law" *Psychological Review* **64** 153-181

von Békésy G, 1949 "The Structure of the Middle Ear and the Hearing of One's Own Voice by Bone Conduction" *The Journal of the Acoustical Society of America* **21** 217

von Békésy G, 1961 "Concerning the Pleasures of Observing, and the Mechanics of the Inner Ear", in *Nobel Lectures, Physiology or Medicine 1942-1962* (Amsterdam: Amsterdam) pp 519-532

von Békésy G, 1967a "Mach Band Type Lateral Inhibition in Different Sense Organs" *J.Gen.Physiol.* **50** 519-532

von Békésy G, 1967b *Sensory Inhibition* (Princeton NJ Princeton University Press)

Wegel R L, Lane C E, 1924 "The Auditory Masking of One Pure Tone by Another and its Probable Relation to the Dynamics of the Inner Ear" *Physical Review* **23** 266-285

Westerkull P, 2002 "BAHA: The Direct Bone Conductor" *Trends in Amplification* **6** 45-52