

Chapter 2

The anatomy and physiology of the mammalian ear

2.1 Introduction

The ear structure is similar across all mammals and interspecies differences appear to be primarily morphological, relating to the size, shape, and layout of the different parts. The mammalian auditory periphery is subdivided into outer (or external) ear, middle ear, and inner (or internal) ear, which includes both the cochlea and the auditory nerve. Subsequently, the auditory nerve projects to the central nervous system, which comprises a number of auditory nuclei in the brainstem, midbrain, thalamus, and cortex.

Much of the experimental and theoretical work in physiological hearing research has revolved around the cochlea, where the most significant signal transformations take place. Whatever kind of transformation or processing occurs in the cochlea, its effects can usually be measured downstream, in the auditory brain. However, since the cochlea is remarkably complex, concealed, and vulnerable, its exact function and micromechanics are not well understood. Other elements of the central part of the auditory system are even less well understood.

As the anatomy and physiology of the ear has been covered in numerous texts in great detail, it will not be covered here in depth. Instead, the emphasis in the following is on coarse-grained components and functions of the human auditory system that are relatively uncontroversial, up-to-date, and informative for a high-level analysis. Therefore, the typical description of the micromechanics of the cochlea or the cell types in the brain with their specific responses is usually avoided. Unless referenced otherwise, the review in §2.2 and §2.4 is largely based on [Pickles \(2012\)](#). Additional presentations of the human ear's structure are found also in [Fuchs \(2010\)](#) and [Rees and Palmer \(2010\)](#) and shorter introductions in [Møller \(2012\)](#) and [Gelfand \(2018\)](#).

When discussing the neurophysiology of the auditory brain, the lack of a coherent theory for it, and for hearing in general, becomes a stumbling block for the understanding of what this complex system does. Therefore, before reviewing the central auditory system, a discussion is provided in §2.3 that provides a few general guiding principles and theoretical ideas that have been commonly employed in auditory neuroscience research, as well as some of their critical shortcomings. The few theories about specific auditory nuclei are then briefly mentioned, where available, throughout the review in §2.4.

Since much of the knowledge about the human ear's anatomy has been gathered from animal data, it is essential to know how the auditory systems of these animal species differ. The final section (§2.5) deals specifically with the comparative anatomy between mammals. It too is a selective and coarse-grained review of some of the notable differences between animals, given the commonalities

covered in §2.2 and §2.4. Therefore, it is somewhat unlike typical comparative hearing texts that tend to focus on specific organs within a subset of animals and to review their similarities as well. This part of the review is based primarily on [Rosowski \(1994\)](#) for the outer and middle ears, [Echteler et al. \(1994\)](#) for the cochlea, and [Glendenning and Masterton \(1998\)](#) for the central pathways. In-depth reviews of various topics in comparative hearing of mammals (and other vertebrates) can be found in volumes edited by [Popper and Fay \(1980, 1994\)](#); [Manley et al. \(2000\)](#) and [Köppl et al. \(2014\)](#).

2.2 The peripheral ear

Many parts of the peripheral ear have at least two anatomical terms that are in regular use. Where available, we present both terms with the lesser used term in the present text given in parentheses.

2.2.1 The outer ear

Acoustic waves in the environment propagate from their source through a medium—either air or water—before entering the external ear, whereupon they begin a sequence of transformations. Sound is diffracted by the **pinna** (or **auricle**)—the cartilaginous external part of the ear, which is connected to the **concha** that is shaped like an irregular horn (Figure 2.1). The pinna and concha collect the impinging sound to efficiently radiate it into the ear.

The sound then passes through a hard-walled **ear canal** (or **auditory meatus**) that terminates with the **eardrum** (**tympanic membrane**), which is a thin membrane that vibrates with the sound pressure. The ear canal has a few resonances that emphasize certain frequencies (most prominently at around 2.5–4 kHz for humans), which are determined mainly by its length ([Wiener and Ross, 1946](#); [Shaw, 1974](#); [Mehrgardt and Mellert, 1977](#)). The sound field that reaches the eardrum is one-dimensional, in very good approximation, especially at low and midrange frequencies, since the ear-canal diameter is much smaller than the wavelength at this range ([Rabbitt and Holmes, 1988](#)). The pinna spectrally shapes the sound as well, but in a direction-dependent manner that has a role in localizing the elevation angle of the sound source (whereas the azimuth is detected using information both ears). This feature, along with its efficient sound power delivery, are thought to be the two main acoustic functions of the outer ear ([Rosowski, 1994](#)).

2.2.2 The middle ear

On the internal side of the eardrum, the vibrations in the air or water medium push a mechanical system made of three small bones, the **ossicles**—the **malleus** (hammer), the **incus** (anvil), and the **stapes** (stirrup)—which constitute together the **ossicular chain** of the middle ear (Figure 2.1). The ossicles are positioned within an **air cavity**, whose pressure is regulated through the **eustachian tube**, which is connected to the pharynx at the back of the mouth. This anatomy was already described in great detail by [Helmholtz \(1945\)](#) (pp. 129–135) along with a hypothetical model of operation that has been partially confirmed since. The malleus is attached to the stretched eardrum in the **umbo** at its center, whereas the stapes is attached via the **annular ligament** to another membrane—the **oval window**, which is part of the cochlear wall. The eardrum, umbo, and ossicles are coupled rigidly at low frequencies, but higher modes of vibration (e.g., bending of the ossicles) and other degrees of freedom in their movement (compression of the ossicular joint or the ligaments between the ossicles, or multimodal motion of the ear drum) may diminish the effectiveness of the coupling at higher frequencies.

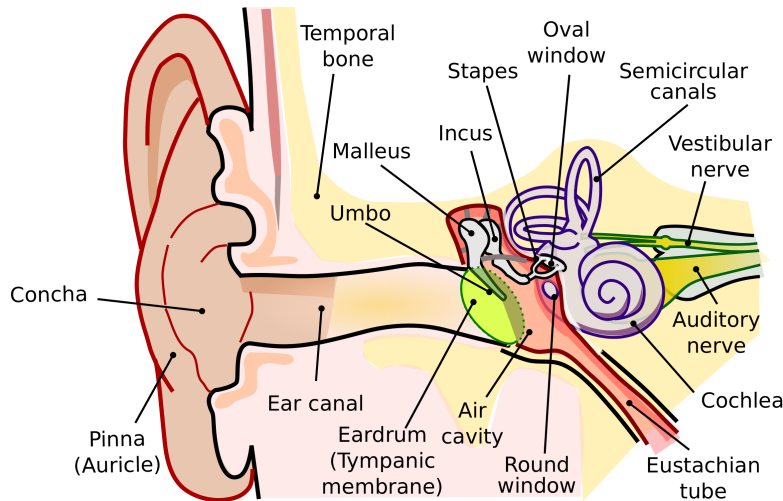


Figure 2.1: The main parts of the peripheral ear (not in scale). The drawing is adapted from https://commons.wikimedia.org/wiki/File:Anatomy_of_the_Human_Ear.svg, which was redrawn after Chittka and Brockmann (2005).

The outer and middle ears work together as a non-ideal (frequency-dependent) transformer between the acoustic impedance (the ratio between the pressure and volume velocity) of the air and that of the incompressible fluid of the cochlea. The middle ear also functions as a mechanical lever that transforms the acoustical power exerted on the eardrum to much larger vibrational power on the oval window, which is of a smaller area than the eardrum. In humans, the gain it provides has a bandpass filter characteristic with a maximum at 1200 Hz and approximately -6 and -7 dB per octave below and above, respectively (Aibara et al., 2001). However, the impedance is not necessarily resistive and complex relations between the pressure and velocity may exist. The vibration of the middle ear is linear with input level below 96 dB SPL and has negligible nonlinear distortion up to 130 dB SPL—well above the ecological sound range (Guinan and Peake, 1967; Aerts and Dirckx, 2010). The middle ear output is the displacement of the stapes that is coupled to the cochlea, which moves one-dimensionally like a piston below 6.7 kHz, but has more complex rocking motion at higher frequencies (Aibara et al., 2001).

Two muscles are connected to the ossicles that control the **acoustic reflex**—the **tensor tympani** is connected to the malleus and the **stapedius** is connected to the stapes. Both appear to regulate the transmission gain of the middle ear, particularly below 1 kHz, by stiffening the ossicular movement, which causes an increase of acoustic impedance. The result may provide some protection against loud sounds and may act as automatic gain control. It is also thought to reduce self-generated sounds by the listeners and to selectively reduce low frequency sounds that can cause undesirable masking. However, the function of these reflexes is not well understood.

The middle ear is a critical stage in hearing that provides necessary amplification in hearing. An impaired middle ear can lead to a **conductive hearing loss**.

2.2.3 The inner ear

The cochlea

The cochlea is a spirally shaped structure that is located inside the anterior portion of the petrous region of the temporal bone, along with the semicircular canals of the vestibular system that occupy its posterior region (Echteler et al., 1994). The coiled tube of the cochlea is longitudinally divided into three fluid-filled compartments, whose cross section is illustrated in Figure 2.2. The

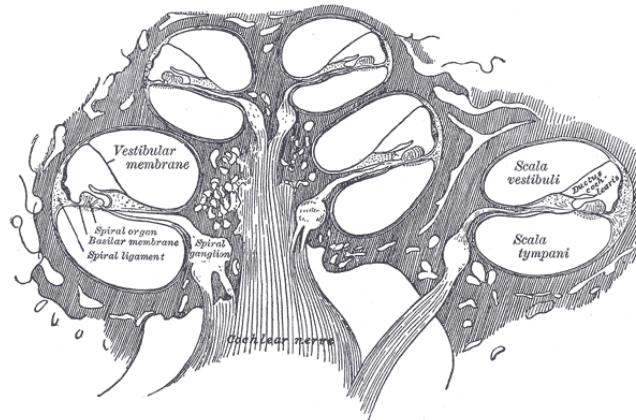


Figure 2.2: A cross section of the in the human cochlea. The narrow sections are apical and the wide ones are basal, showing the typical 2.5–3 turns of the human cochlea (Pietsch et al., 2017). The helicotrema is found at the top of the cochlea (not shown). Image by Henry Gray from the “Anatomy of the Human Body” (1918), taken from <https://www.bartleby.com/107/illus928.html>.

first compartment, the **scala vestibuli**, is where the oval window is located. The motion of the stapes presses the elastic oval window and thus the cochlear fluid, the **perilymph**, which is largely incompressible (with approximate fluid properties of water and composition similar to extracellular fluid). Scala vestibuli terminates in an opening called the **helicotrema**, where the compartment connects to **scala tympani**, which is another compartment that leads back through the entire length of the cochlea. The oval window displacement causes a corresponding displacement of the **round window**, at the other end of scala tympani.

The two scalae are separated by elastic membranes, which tightly enclose on the **scala media** (or **cochlear duct**)—another compartment that is not directly connected to the other two and is filled with **endolymph**, or **endocochlear fluid** (see Figure 2.3). The endolymph is similar in properties to intracellular fluid that is rich with potassium and low in sodium ions that make the scala positively charged in reference to the perilymph by about about 80–100 mV. The supply of ions originates in the **stria vascularis**, on the wall of scala media, which has a capillary supply. The membrane below scala vestibuli is called **Reissner’s membrane** and it does not have a direct role in the sound transduction chain, although it has been proposed that it takes part in the reverse transmission the otoacoustic response from the ear (Reichenbach et al., 2012). The fibrous membrane above scala tympani is called the **basilar membrane** (BM) and it is attached to the **organ of Corti**—a complex structure of cells that transduces the mechanical movement to electric discharges in the auditory nerve. The elastic BM moves in response to a pressure gradient between the scala vestibuli and scala tympani. This movement manifests as a slow **traveling wave** that propagates from the oval window in the **base** of the cochlea toward the helicotrema in its **apex**.

The cochlea has unique frequency analysis properties, whose mechanics was largely elucidated by Békésy (1960). High frequencies conducted through vibrations in the stapes cause the BM to vibrate at the base, while low frequencies vibrate at the apex, in terms of the traveling wave peak response. For a given stimulus frequency, the wavelength and phase of the traveling wave change quickly and slows down just before the frequency-dependent peak, whereupon it dies out almost immediately after the peak with little to no energy reflected back to the base. The peak sharpness is reduced at high intensities. The frequency analysis property is achieved by the particular geometry and mechanics of the cochlea, which is tapered in two opposite directions: the ducts are wide at the base and narrow at the apex, whereas the BM is wide at the apex (about 0.5 mm) and narrow at the base (0.1 mm). Critically, the forward direction of the movement is largely determined by the

stiffness of the BM that is high at the base and gradually diminishes towards the apex, where the BM movement is mass-controlled by the perilymph. The peak sharpness is limited by damping of the cochlear partition (the organ of Corti including the BM).

The organ of Corti is attached to the BM by supporting cells that embed the **inner hair cells** (IHCs) and **outer hair cells** (OHCs). Both hair-cell types comprise of a **soma** (cell body) and a **hair bundle**, or **stereocilia**. These stiff hair-like structures protrude from the apex of the cell, which is flushed with the **reticular lamina**—a thin membrane inside the organ of Corti, which chemically isolates the perilymph from the endolymph. The IHCs are organized in a single row and the OHCs in three or more rows along the cochlea. The hair bundle is graduated in height and it forms straight (IHCs) or V-shaped (OHCs) patterns that are apparent on the reticular lamina. The stereocilia within one hair cell are connected to one another via small **horizontal top connectors** and **tip links**, so that the entire hair bundle rigidly deflects upon mechanical force and it bends around its base. The **side links** of the OHCs provide mechanical reinforcement against loud sounds (Han et al., 2020).

The traveling wave in the BM causes a shearing movement of the organ of Corti, which leads to deflection of the stereocilia of the hair cells. In turn, this deflection of the IHCs causes the tiny tip links between the hairs to open and close mechanotransducer channels in the apical end of the stereocilia, close to the tip links. This leads to ionic current flow from the positively-charged endolymph to the negatively-charged cell and then to neurotransmitter release (likely, glutamate, although not exclusively; Eybalin, 1993) to the synapse of the auditory nerve. The opening of the channel is asymmetrical with respect to the direction of deflection of the hairs, which results in a mixed AC and DC response potential. The hair cell response reflects the mechanical response of the BM around its peak, whose corresponding frequency is referred to as the **characteristic frequency** (CF)¹⁹.

The **tectorial membrane** (TM) is an elastic gelatinous membrane that is flapped over the hair bundle rows and the reticular lamina. Unlike the BM, the TM is found in all tetrapods, but its role in the cochlear micromechanics remains uncertain. The tallest OHC stereocilia are connected to one another and to the TM by **TM-attachment crown**, which is essential for the normal OHC function (Han et al., 2020). It has been recently found in guinea pigs that both the IHCs and OHCs are deeply embedded inside the TM (Hakizimana and Fridberger, 2021). It was also found that all hair cells are connected to the TM with filamentous tubes called **Ca²⁺ ducts**, which directly supply calcium ions to the hair cells that are required for the OHC motility. In the same study it was shown that the TM rests on the RL with no gap between them in quiet, and that their movement remains tight also at large amplitudes, only with momentary gaps when the system returns to quiet. During stimulation, the OHCs and the IHCs appear to be phase locked. These results contradict past findings that were done using coarser methods that found that the only Ca²⁺ supply to the bundle is from the endolymph, that only the tallest OHC tips are embedded in the TM, and that the IHCs are not directly connected to it, although their tips fit into a groove on the bottom of the TM (**Hensen's stripe**).

It was demonstrated in human cadavers and in mice that the TM sustains a traveling wave as well, which in comparison with the BM has sharper tuning, apically shifted frequency mapping, and larger dynamic range (Lee et al., 2015; Farrahi et al., 2016). Shearing motion of the RL translates to TM motion that can move the IHC hair bundle through fluid forces (Zwislocki, 1980; Cheatham

¹⁹The majority of the classical studies of the cochlear mechanics in vivo targeted basal CF sites, whose responses were later extrapolated to apical (low-frequency) sites. Recent studies have begun challenging the validity of this extrapolation, as they consistently revealed lower frequency selectivity at the apex, which does not correspond with a bandpass model. The interaction between the filtering of the cochlea and the auditory nerve is likely to be different there as well. See §13.4.1 for a short review and relevant references.

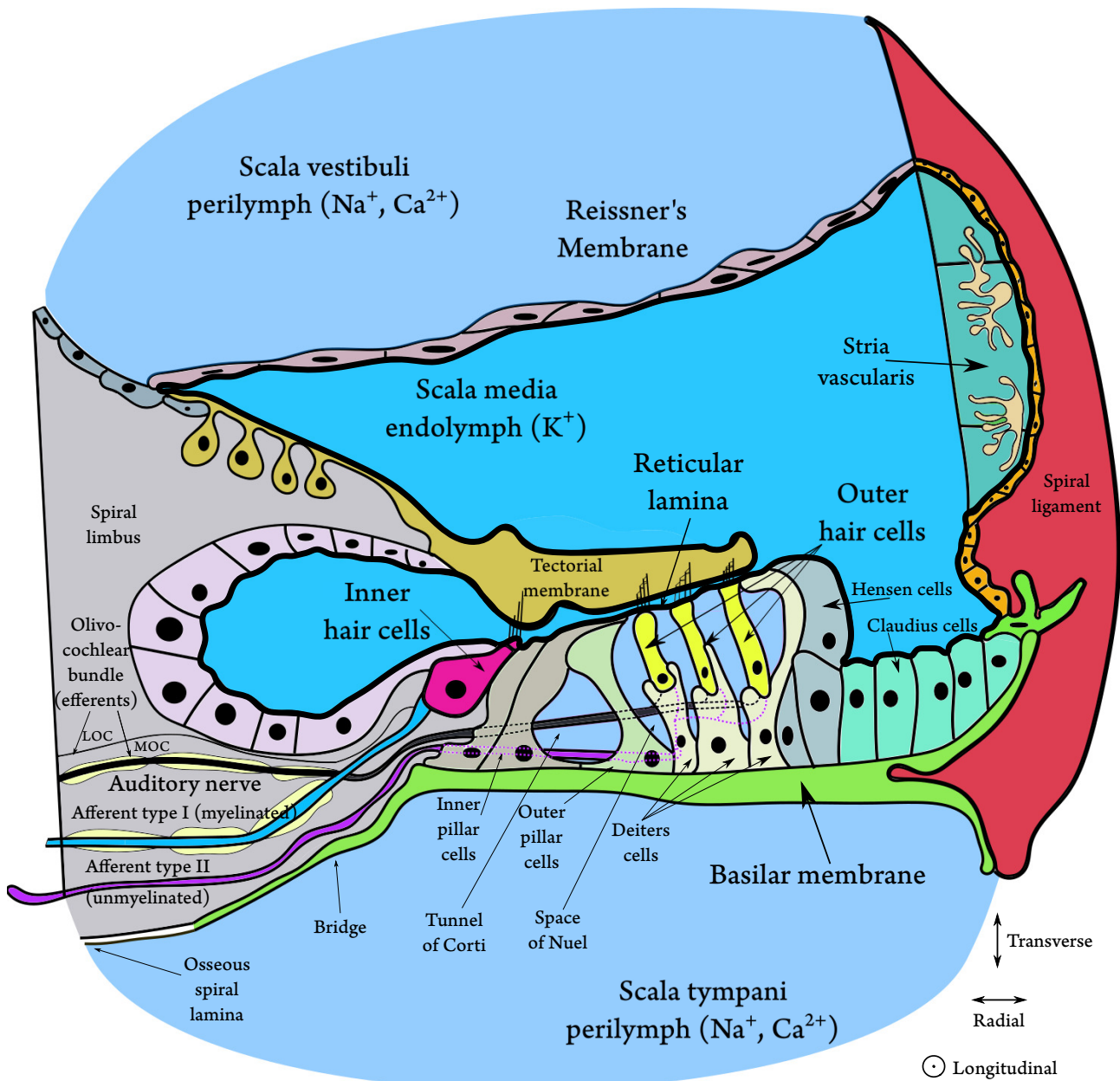


Figure 2.3: A cross section of the organ of Corti, showing its most important elements. The figure is based on Figures 2.2, 2.3, 2.13, and 2.21 in [Slepecky \(1996\)](#) and on Figure 3.5 in [Pickles \(2012\)](#). The thick black line surrounding the scala media and the endolymph represents the tight junctions that electrically separate it from the perilymph and mark the boundaries of the cochlear duct. New evidence has shown that the stereocilia of both hair cell types are deeply embedded in the tectorial membrane (TM) and that there is no gap between it and the reticular lamina without acoustic stimulation ([Hakizimana and Fridberger, 2021](#)). There is a gap of endolymph illustrated, as a halfway reminder of older observations that showed free standing IHCs under the TM. The bridge tissue is after [Raufer et al. \(2019\)](#).

and Dallos, 1999; Robles and Ruggero, 2001). Other recent findings in mice suggest that the role of the TM may be in stabilizing the cochlear mechanical amplification (see below), as detached TMs result in high incidence of spontaneous otoacoustic emissions (Cheatham et al., 2016).

The passive system in the cochlea cannot achieve on its own the tuning sharpness of the BM that is observed empirically, so there is little doubt today that the OHCs provide the compressive amplification necessary for the healthy cochlea to hear sounds at adequate levels (Robles and Ruggero, 2001). The major candidate mechanism that can drive the amplification is somatic motility, but another active mechanism exists in the OHCs—the bundle motility—whose potential role in amplification has not been clearly elucidated. Both mechanisms are induced by the shear forces on the organ of Corti, which moves with the vibrations of the traveling wave in the BM. This cochlear motion causes deflection of the stereocilia, which gates mechanotransducer ion channels that set the current flow into the cells.

The main active mechanism proposed for amplification—and the most touted one—is the **somatic motility**, or **electromotility**, of the hair cells. It involves length changes of the hairs when intracellular electric current passes through the cells and causes somatic lengthening with hyperpolarization and shortening with depolarization (Brownell et al., 1985). In vitro, this movement can take place at ultrasonic frequencies (shown at least up to 79 kHz, Frank et al., 1999). However, the exact dynamics of this mechanism—how it actually delivers power that impacts the signal picked up by the IHCs—is not entirely clear. One influential class of amplification models requires that the motile changes happen in phase with the traveling wave, on a cycle-by-cycle basis (Dallos et al., 2008; Guinan Jr, 2020). However, this view has been challenged because electromotility may be limited to much lower frequencies by the capacitive cell membrane in vivo and more realistic in vitro conditions (Dallos and Evans, 1995; Ashmore, 2008; Vavakou et al., 2019; Santos-Sacchi, 2019; Santos-Sacchi and Tan, 2019). Electromotility is associated with the protein prestin (Zheng et al., 2000), and in mutated mice that lack prestin, there is no amplification or compression (Lieberman et al., 2002). Counterexamples exist for the presence of prestin with no amplification in birds (e.g., in chickens, Xia et al., 2016), and electromotility on its own does not guarantee amplification also in mammals (gerbil in Strimbu et al., 2020). Also, amplification with no prestin was observed in prestin-knock-out mice at ultrahigh frequencies (> 40 kHz) (Li et al., 2022). Alternative models to electromotility-based amplification work by locally modifying the impedance of the basilar membrane, thereby creating a sharper resonance with less dampening (e.g., Kolston, 2000; Vavakou et al., 2019). These length changes, as well as other biophysical changes in the cell, can result in substantial OHC stiffness modulation (Ashmore, 2008).

The OHCs have an additional active part, which is the hair bundle (the stereocilia), whose movement is called **bundle motility**. The hair bundle protrudes from the apex of every hair cell, at the reticular lamina, and is most likely mechanically connected to the TM. It has been proposed that this motion that happens during depolarization of the OHCs can impart power to the motion of the reticular lamina and, therefore, may provide amplification as well (Kennedy et al., 2006). However, much evidence to support this mechanism is still lacking.

It seems that whichever mechanism is responsible for the cochlear amplification, it requires a certain feedback between the BM, the OHC soma, the hair bundle, and perhaps other elements such as the TM (see §9.8.2 for feedback model references). It was also hypothesized that the two motile mechanisms may be both necessary to achieve the OHC amplification in mammals (Peng and Ricci, 2011).

The OHCs are implicated in a host of nonlinear phenomena in the cochlea. First and foremost, they provide amplification around the characteristic frequency for low-level inputs, which results in effective lowering of the hearing threshold. Consequently, it also sharpens and shapes the filter response around the CF. Additionally, the OHCs cause compression of the dynamic range of the input

at medium levels (approximately 30–90 dB SPL), they are implicated in intermodulation distortion generation, in two-tone suppression, and in otoacoustic emissions. The OHCs require metabolic supply of energy, which means that their associated functions cease to work after death.

Recent evidence has suggested that, the amplification and the other active effects of the OHCs take place inside the organ of Corti, so they are measurable on the reticular lamina, but not necessarily in the BM (e.g., [Ren et al., 2016a](#); [Cooper et al., 2018](#); [Nuttall et al., 2018](#); [He and Ren, 2021](#); [Lin et al., 2024](#)). Furthermore, the stiffness and ionic regulation that are observed in the response of the OHCs are regulated by the supporting cells, whose role is gradually being uncovered ([Lukashkina et al., 2022](#); [Zhou et al., 2022](#)). It was suggested that the force generated by the OHCs may in fact be directed toward the TM and the reticular lamina, and hence the entire Organ of Corti, which then indirectly applies pressure that amplifies the traveling wave, rather than by forcing the BM directly ([Altoè et al., 2022](#); see also [Guinan Jr, 2022](#)).

The auditory nerve

The vestibulocochlear nerve, which is the eighth cranial nerve, is shared between the vestibular and auditory (cochlear) nerves ([Rea, 2014](#), pp. 81–93). The human cochlea is innervated by approximately 30000 afferents of the **spiral ganglion** neuron type and about 1400 efferents, which form the **auditory nerve**. Most afferent fibers are **Type I**, which are fast (both fibers and soma are myelinated) bipolar cells, whose dendrites synapse to IHCs and axons synapse to the auditory brainstem nuclei. Each IHC is connected to 10–30 Type I fibers with a single **ribbon synapse** per fiber, which are characterized by their reliable, temporally precise, and sustained responses. **Type II** fibers are slow (unmyelinated) unipolar cells that innervate up to 50 OHCs each. The OHCs are synapsed to multiple Type II fibers. Both types project to the cochlear nucleus in the brainstem, which is the first nucleus in the central auditory pathway.

It is common to refer to the discharge patterns in the auditory nerve as the earliest part of an auditory **neural code**. It has been extensively studied in single-unit recordings in many species, mostly for relatively simple signals (e.g., [Eggermont, 2001](#); [Rutherford et al., 2021](#)). The code varies in its rate and inter-spike patterns, with respect to the input stimulus level, temporal, and spectral characteristics. Afferent fibers (Type I) are characterized by their spontaneous discharge rate, which reflects their inherent noise level and may be low, medium, or high. High-spontaneous-rate fibers are most sensitive to low-level inputs and are the most prevalent (90% of all fibers), but have a relatively limited dynamic range of about 20–30 dB. The low-spontaneous-rate fibers are sensitive to high levels and have wider dynamic range of up to 60 dB. The frequency tuning of the fibers reflects the bandpass characteristics of the cochlea, which is often illustrated using **tuning curves** of the effective filtered response. At the CF, the spiking rate is maximal (for a given input level). Another important feature of the auditory nerve spiking is **phase locking**—a precise temporal correspondence between the (tonal) stimulus wave phase and the neural spiking pattern in the auditory nerve and most other nuclei. Phase locking is limited to low frequencies (estimated to be below 4 kHz for humans). The auditory nerve can also track the low-frequency stimulus envelope for all carrier frequencies (also above 4 kHz), which is also conserved throughout all auditory nuclei further downstream to the auditory cortex ([Souffi et al., 2023](#)).

The function of Type II fibers has been recently elucidated in experiments on mice *in vivo*, where a non-acoustic nociceptive (pain) reaction to very loud sounds (120 dB SPL) has been established, which leads to avoidance behavior ([Flores et al., 2015](#)). *In-vitro* tests found that these fibers react to OHC damage, as the kind that occurs with the traumatic exposure to loud sounds and is irreparable in mammals ([Liu et al., 2015](#)).

Descending efferent innervation to the hair cells exists in the cochlea as well, which is called the **olivocochlear bundle**. It arises in the superior olivary complex (SOC) of the brainstem and

has two divisions. The medial olivocochlear (MOC) efferents, are myelinated and they project both to the contralateral and ipsilateral OHCs with cholinergic synaptic contacts to the cell body. The lateral olivocochlear (LOC) fibers are unmyelinated and they project mainly to the ipsilateral afferent dendrites of the IHCs, but not directly to the hair cell body. The MOC receives afferent input from the auditory nerve via the poteroventral cochlear nucleus (PVCN), which projects to the contralateral SOC. Activation of the MOC causes a bilateral change of threshold through the **medial olivocochlear reflex**, which has two pathways—an ipsilateral reflex and a contralateral reflex (Liberman and Brown, 1986; Brown, 1989). Fibers that respond to monaural sounds are the majority and a much smaller number of binaural fibers respond to stimulation from both ears. Both ipsilateral and contralateral fibers synapse to the OHCs in the same way, although there are about double the number of ipsilateral as there are contralateral. Accordingly, the ipsilateral reflex effect has been consistently stronger than the contralateral effect in humans (Salloom and Strickland, 2021). The MOC efferents are tuned to the same frequencies as the OHCs they innervate, although their activation is proportional to the bandwidth of the stimulus, where broadband noise is typically used. While uncertain, it has been proposed that the MOC may provide some protection against acoustic trauma including own voice attenuation, regulate the operating point of the amplification, optimize detection of signal in noise (i.e., reduce masked tone threshold), and reduce the sensitivity to unattended stimuli (see also § 16.4.2). The role of the LOC remains unknown at present. See Romero and Trussell (2021) for a more detailed neurophysiological account of the olivocochlear system.

2.3 Organizing principles and common threads in the central auditory system

Unlike many parts of the peripheral ear, the functions of the different auditory nuclei have so far eluded a plain explanation. The auditory neuroanatomy has been thoroughly charted, but the complexity of the parts forming the system as a whole is prohibitive for the formation of a simple functional account. In order to facilitate the description of the system, several general and particular aspects of the auditory brain are discussed in brief. Some of these principles are not unique to hearing.

The central auditory system comprises several nuclei and multiple pathways for the signal to travel downstream to the cortex (the major pathways are illustrated in Figure 2.4). Each nucleus has several neuron types, which can be characterized by different morphologies and typical responses to different stimuli, usually using single-unit (e.g, one neuron) measurements. The auditory nuclei further project to other auditory and occasionally non-auditory nuclei, which may be either excitatory or inhibitory. The connections and responses of the different neuron types have been studied in depth over the last decades, and functions of the respective areas were sometimes inferred from these responses. However, the complexity and diversity of the complete system makes it very challenging to attribute a “closed-form” function to most, if not all, of the auditory nuclei.

2.3.1 Tonotopy

Perhaps the most characteristic organizing feature of the auditory system is that the cochlear frequency axis, its **tonotopy**, is conserved throughout the auditory brain. Thus, the same monotonic arrangement of frequencies from high to low is found in all of the main auditory nuclei. For example, in the cat, tonotopy was recorded in the cochlear nucleus, lateral lemniscus, inferior colliculus (IC), medial geniculate body (MGB), and auditory cortex (Bourk et al., 1981; Aitkin et al., 1970;

Merzenich and Reid, 1974; Aitkin and Webster, 1972; Reale and Imig, 1980). Importantly, even though the early brainstem nuclei have their own tonotopic maps, they all converge to a single map in the inferior colliculus (Casseday and Covey, 1996). Additionally, in the human primary auditory cortex, tonotopy appears to be more complex than in the more peripheral nuclei, as it maps to pitch rather than frequency, as was initially demonstrated measured in the magnetoencephalogram (MEG) response to the missing fundamental (Pantev et al., 1989). The most recent measurements using functional magnetic resonance imaging (fMRI) suggest that frequency and pitch are both topographically mapped in the auditory cortex by different neural populations that have only limited overlap (Allen et al., 2022).

Starting from the IC and extending to the MGB and auditory cortex, a useful classification of the different auditory stations is into **lemniscal pathways** that are more sharply tuned and are organized tonotopically. They are the primary ascending conduit of auditory sound information from the periphery (Carbajal and Malmierca, 2018). In contrast, **non-lemniscal pathways** constitute the “belt” areas that are less responsive to auditory stimuli, are not organized tonotopically, while they send ascending pathways to the next nucleus, and receive descending connections from the cortex. However, it has been recently suggested that the non-lemniscal corticocortical circuits may be involved in complex processing of speech that is parallel to that of the core and its associated lemniscal pathway (Hamilton et al., 2021). We will mostly consider the lemniscal nuclei in the review below.

2.3.2 Synchronized responses

Another important property that is used to characterize different neuron types is their degree of phase locking, or synchronization—how well they track the acoustic waveform of the stimulus (see §9.7.2). Some nuclei and particular cell types seem to excel in synchronizing either to the carrier or to the envelope. The degree of synchronization to the stimulus can be enhanced through an increase in the number of dendrites that synapse the synchronizing cell. This is the case since multipolar cells generally have high threshold for firing, which means that several input fibers have to fire simultaneously in order for the cell to fire. When all the inputs are excitatory, such a configuration is sometimes referred to as a **coincidence detector**—a useful circuit in the enhancement of the temporal response of the neuron (see §8.5). Another aspect in which cells vary is whether they synchronize to stimulus onsets or to sustained stimuli, as some cells synchronize to one feature and not to the other. Either way, the response usually shows spiking adaptation over time—a reduction in the average spiking rate—in response to an unvarying stimulus.

2.3.3 Generalizations from single unit recordings

Because of the complexity of the auditory system, as well as neural systems in general, it is difficult to attribute specific functions for the brainstem circuitry. Hence, the understanding of the roles of many auditory nuclei remains relatively vague. Pickles (2012, pp. 155–157) categorizes the central auditory system research findings into overarching themes, all of which are also encountered in non-auditory sensory research²⁰:

1. Feature detection and extraction—Relevant auditory features from the stimulus (e.g, spectrum, temporal fluctuations, sound location) may be observable through recordings of a single unit. It may also be associated with a population of neurons that is localized in a certain brain area and has a modular role in the complete signal processing chain.

²⁰Pickles (2012) presented three themes, but we merged the first two into a single theme, as they are difficult to distinguish in the original text.

2. Hierarchical analysis—It is sometimes possible to demonstrate progression in complex auditory processing that emerges in one auditory area and culminates in another, usually more central in the ascending pathways (e.g., the auditory cortex as the destination area of auditory scene analysis). This theme introduces continuity and causality into the information-processing logic of different auditory areas in the brain.

These themes are interdependent and find wide use in auditory research. However, they require external theories to direct the experimental explorations and inspire plausible interpretations. A lack of theory makes the interpretation much more challenging due to the complexity of the system. This was illustrated in a critical paper called “*Could a neuroscientist understand a microprocessor?*”²¹ (Jonas and Kording, 2017). The premise of the paper is that reverse engineering of even a relatively simple computational circuit may be downright impossible using many of the standard methods in neuroscience. Their example was a (simulation of a) popular microprocessor from 1975 that contains 3510 transistors in total, which in today’s standards would be considered primitive. It was made to run a number of simple video games as known “behaviors”. Transistors in the circuit were taken as rough analogs of neurons. Employing classical neuroscientific methods produced a wealth of data and many distinct patterns, which were nevertheless useless in explaining what the microprocessor actually does on a level that can be generalized beyond the specific game being played.

While these conclusions from Jonas and Kording (2017) may appear dispiriting (and perhaps controversial), they are invoked to highlight the nontrivial implications of a lack of a coherent theory of the auditory brain pathways. Local simplicity of brain functions, including low-level ones as observed using simple stimuli in single unit recordings, might be misleading, especially if tested with a narrow range of stimuli. As illustration, a recent comparison of several phenomenological and biologically-inspired auditory signal processing models demonstrated that the simplest model has the highest correspondence to single-unit recordings from the ferret primary auditory cortex, using several simple stimuli (clicks, pure tones, white noise, speech snippets; Rahman et al., 2020). The authors suggested that this surprising result can entail that the total effect of the auditory system is much simpler than its complexity may imply. However, their interpretation entails (with some exaggeration, admittedly) that for the tested stimuli the brainstem processing may as well be replaced with direct connections to the cortex, as long as the necessary frequency weighting and nonlinear compression is reproduced. While not unique to this model or study, such oversimplification leads to tagging entire networks with the infamous “relay neuron” role (often designated to the thalamus). To this Winer and Schreiner (2005b, p. 46) commented: “*The concept of a relay nucleus requires critical scrutiny as all central nuclei transform and modify the information that passes through them...*” Another problem with this interpretation is that it implies that the single-unit primary auditory cortical response is tantamount to a perceptual and behavioral output. It was recently argued that spiking patterns cannot be simply replaced in analysis with more informative signals or inputs, as these patterns do not generalize to real-world stimuli with arbitrary context (Brette, 2019).

Another point of view to consider continues the microprocessor analogy, where the auditory brain is thought to perform some computational task. Hypothetically, any computation requires not only the transmission of the data, but also the transmission of various control signals to modulate its processing, according to the situation. The auditory system processes signals that are on the same temporal scale as the neural processing speed limit. Thus, neuron spiking patterns may correlate to acoustic signals with relatively few transformations throughout the brain, which reduce its complexity and maintains clear correspondence between stimulus and response. However, internal control signals that may be internally used for computational purposes may be correlated with the stimulus as well.

²¹It was inspired by an earlier paper called “*Can a biologist fix a radio?—Or, what I learned while studying apoptosis*” (Lazebnik, 2002), which may have some relevance to complex structures in the peripheral ear, such as the organ of Corti.

This may result in a cacophony of control and sensory signals in the brain that are correlated on a population level, but have fundamentally different roles in the system as a whole.

In summary, getting a handle on the brain function requires a theory for guidance, which is especially pertinent in the auditory system due to its high complexity. Developing a theory from the reduced low-level components may be hypothetically possible, but it will require a way about compressing the amount of low-level details to a compact description, which can be formulated using high-level concepts. There is no guarantee that this is the case, though. Quoting [Winer and Schreiner \(2005b\)](#), p. 45) again: *“It may not be possible or even appropriate to delegate functions to nuclei, as functions are global constructs, while neurons and nuclei and circuits are restricted to local operations.”*

2.3.4 Dual-stream models

We shall mention one organizational theory that has been influential in recent hearing research that may apply to the entire auditory system—the **dual stream model**. It was originally proposed for vision by [Trevarthen \(1968\)](#) and [Schneider \(1969\)](#) and has been revised several times since. According to this model, the processing of visual objects bifurcates in the cortex. The **ventral stream**, or the **what stream**, processes the pattern (shape) and the identity of the object, whereas the **dorsal stream**, or the **where stream**, processes its position in space ([Mishkin et al., 1983](#); [Wilson et al., 1993](#)). Analogous functions were identified in two distinct anatomical streams going out of the auditory cortex that pertain to identification and localization ([Rauschecker, 1998](#); [Kaas and Hackett, 1999](#); [Romanski et al., 1999](#)).

The neat labor-division offered by the dual-stream model turned out to be more complex both in vision and in hearing. In vision, object localization processing is tied to motor control functions that may be required for actions based on the visual input ([Goodale and Milner, 1992](#)). Thus, a revised dual-stream model in vision distinguishes vision-for-perception (ventral) and vision-for-action (dorsal) streams ([Goodale, 2011](#)). A more recent version of the model identifies a third visual processing stream that specializes in motion or socially relevant objects (e.g., faces, body movements) ([Pitcher and Ungerleider, 2020](#)). A similar refinement has been applied in hearing, as it became clear that the dorsal stream does not process localization exclusively, but also has some roles in speech and even music processing that require tracking acoustic changes over time ([Belin and Zatorre, 2000](#); [Hickok and Poeppel, 2004, 2007](#)). The ventral stream in speech may be suitable for recognition of words, which are sometimes classified as auditory objects. Thus, a unified model is emerging that posits that language may require processing that relies on both streams, which include sensorimotor connections with articulators that may be used to produce speech ([Rauschecker, 2017, 2018](#)). This interpretation would be a special feature of the human cortex, with unknown applicability to other mammals.

The dual-stream model has also influenced the interpretation of subcortical processing that is known to bifurcate early in processing (in the cochlear nucleus) into several parallel streams that converge again only in the midbrain (inferior colliculus; [Pickles, 2012](#), Chapter 6). The early ventral stream does seem to be dedicated mostly to binaural localization, but the roles of the other two streams are not obvious. Another version of the dual-stream model for speech suggested that the brainstem may be separately processing the fundamental frequency of speech and spectral changes in time—at least as early as the lateral lemniscus and inferior colliculus ([Kraus and Nicol, 2005](#)). However, as has turned out in the cortical stream modeling, the subcortical streams do not readily lend themselves to neat classification as the original what/where model suggested, or to other obvious signal processing, so their explicit functions are not necessarily clarified by this model.

2.4 Central auditory neuroanatomy

This section provides a simplified overview of the complex circuitry in the auditory brainstem. Only high-level details that are deemed to have greater functional significance for the complete system are mentioned, while most of the fine-grained details are omitted.

2.4.1 Medulla and pons

Auditory nerve fibers innervate the cochlear nucleus, where they branch to three **cochlear nucleus** (CN) areas. The axons from these nuclei run in three different tracts, or **acoustic strias**, which lead to further nuclei ([Cant and Benson, 2003](#)). The rostral branch begins from the **anteroventral cochlear nucleus** (AVCN), which is one of the two divisions of the **ventral cochlear nucleus** (VCN)—the other being the **posteroventral cochlear nucleus** (PVCN). The AVCN receives its input from the **endbulbs of Held** — giant synaptic terminals with multiple synapses. The endbulbs of Held are characterized by reliable spiking in response to auditory nerve spikes, as well as very short delays as the synapses are positioned on the soma itself. The AVCN appears to have little or no inhibitory inputs, so its response largely reflects that of the auditory nerve.

The AVCN projects contralaterally to the **medial superior olivary** (MSO) and ipsilaterally to the **lateral superior olivary** (LSO)—both are nuclei of the **superior olivary complex** (SOC). The LSO also receives inhibitory input from the contralateral VCN via the **medial nucleus of the trapezoid body** (MNTB), which contains the largest synaptic terminals in the brain—the **calyx of Held**—that enable temporally precise spiking. The SOC is critical in binaural functions, where the LSO primarily extracts interaural loudness difference (ILD) cues at high frequencies, while the medial superior olive (MSO) primarily extracts interaural time difference (ITD) cues at low frequencies. The MSO has additional capabilities in extracting fine temporal cues that are monaural (e.g., echo suppression), at least in small mammals that do not exploit interaural localization cues, such as bats and rats ([Grothe, 2000](#)). The SOC is also where the olivocochlear bundle arises, which projects contralaterally to the OHCs (the MOC), ipsilaterally to the IHCs (the LOC), as well as collaterally to the CN.

The caudal branch of the CN innervates the PVCN and the **dorsal cochlear nucleus** (DCN). The PVCN contains four different cell types with unique responses. Among them are onset detection by **octopus cells** of broadband sound (e.g., clicks), which are temporally precise. Another interesting cell type is the **chopper** (**T-stellate**) multipolar neuron, which is sharply tuned in frequency and fires with a sustained period that is independent of the input frequency but closely tracks its envelope. The T-stellate cells project to the MOC cells in the SOC, among others. The PVCN cells project ipsilaterally and contralaterally to most neighboring nuclei, including the VCN, the ventral nucleus of the lateral lemniscus (VNLL), and the **inferior colliculus** (IC). Additionally, they project to the **superior paraolivary nucleus** (SPN), which is a part of the SOC, but has a monaural function in precise offset and (much less in) onset detection in complex sounds ([Felix II et al., 2017](#); [Kopp-Scheinflug et al., 2018](#)). Therefore the PVCN plays a role in all auditory functions, yet a single specific role cannot be easily pinned down.

The DCN is composed of a complex network of cells with diverse responses and numerous connections to other nuclei. In addition to auditory inputs, it receives and projects to somatosensory cells around the pinna. This is thought to reflexively affect localization optimization in animals that can move their external ears, which can have a role in detecting elevation localization cues. It may also have a role in suppression of the input of own vocalizations. Like other areas in the CN, the DCN receives excitatory and inhibitory inputs from most other auditory nuclei, which can change the tuning curve bandwidth of the different cells, or modify their dynamic range by shifting the

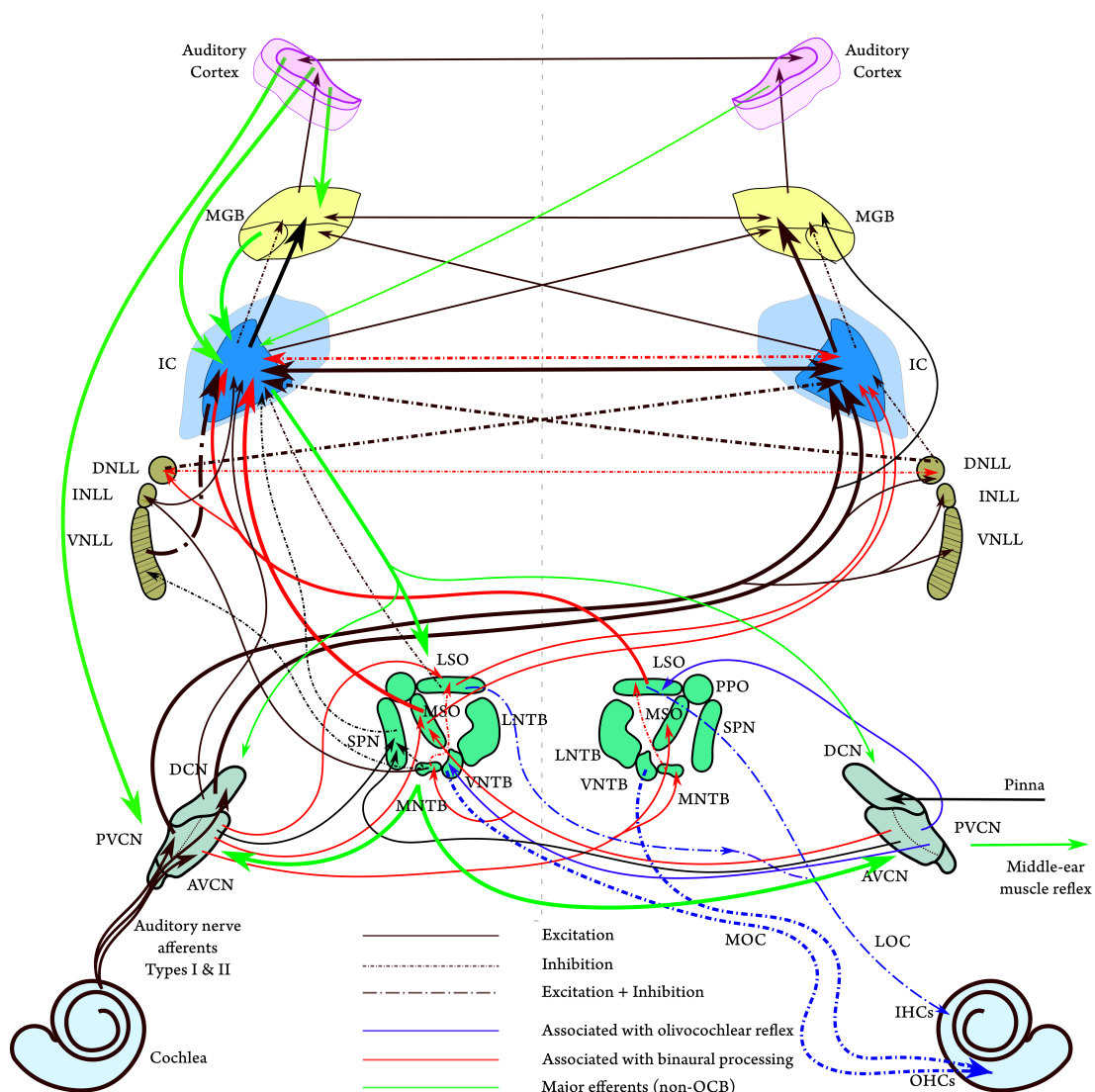


Figure 2.4: A schematic diagram of the main connections in the auditory brain, with emphasis on major afferents (in black) up to the central nucleus of the inferior colliculus (IC). Note that some projections are united to avoid graphical clutter, which may not reflect the actual anatomy. The most prominent projections are plotted with thicker lines. The most important efferent connections without specific subnucleus destination are plotted in green, based on Figures 3.2–3.5 in Schofield (2010). The olivocochlear bundle (OCB) is displayed in blue along with the afferent projections that inform it according to Figure 1 in Lopez-Poveda (2018). Excitatory projections are in solid lines, inhibitory in short-dash-dot, and mixed inhibitory-excitatory in long-dash-dot (see Larsen and Liberman, 2010 for a specific discussion of the lateral olivocochlear, LOC). The main structure is plotted after Figure 6.12 in Pickles (2012), Figure 1A in Felix II et al. (2018), and Figure 2.10 along with additional information from Malmierca and Hackett (2010). The main projections that are associated with binaural processing are plotted in red after Figures 6A and 8D in Grothe et al. (2010). The contour of the cochlear nuclei (CN) in humans was plotted after Figure 7 in Moore and Osen (1979). The superior olivary complex (SOC) contour of humans is plotted after the micrograph of Figure 2A in Weinrich et al. (2018). The nuclei of the lateral lemniscus (DNLL, INLL, and VNLL) are plotted after that of the cat in Figure 2 in Glendenning et al. (1981) and Figure 11.11 in Langner (2015). The human inferior colliculus (IC) contour was plotted after Mansour et al. (2019, Figure 3). The human medial geniculate body (MGB) was plotted after Winer (1984, Figure 1). The auditory cortex was plotted after Kaas and Hackett (2000, Figure 2B). Connections to and from the IC, MGB, and auditory cortex do not imply specific target subnuclei. Note that the direct projection from the CN to the MGB is plotted as if coming from the VCN, but some studies pointed to the DCN as the origin of this pathway (Schofield et al., 2014). For a complete map of known projections from the CN see Cant and Benson (2003). A diagram of the interneuron connections within the CN is provided in Young and Oertel (2018). Nonauditory connections to the IC are reviewed in Gruters and Groh (2012). Connections between the CN, IC, A1, and A2 and the cerebellum are discussed in Mennink et al. (2020).

threshold for firing. Its main projections are to the inferior colliculi on both sides. The DCN is also implicated in sound offset detection and as such it may constitute a part of a specialized “offset pathway” along with the SPN, the IC and the medial geniculate body, which appears to be distinct from an “onset pathway” in the auditory brain (Kopp-Scheinflug et al., 2018).

The CN receives centrifugal (descending) efferent innervation from the IC and from the auditory cortex, which is implicated in modulation of the critical bandwidth of the auditory filters and changing of the masked threshold for tone in noise. The centrifugal inputs may also affect the time constants associated with masking, as well as spike timing modulation by somatosensory inputs to DCN neurons.

The next stage after the CN and SOC is the **lateral lemniscus tract**, which has three nuclei—the **ventral, intermediate, and dorsal nuclei of the lateral lemniscus** (VNLL, INLL, and DNLL, respectively). The DNLL is part of the localization stream as it receives inputs from the LSO, MSO and CN and projects to the IC on both sides. It has been suggested that the DNLL is key in inhibiting the reflection (the lag) in the (binaural) precedence effect, which then gives precedence to the direct (the lead) sound at the level of the IC (Brown et al., 2015).

The VNLL receives projections from the ipsilateral MNTB and the contralateral VCN, but is completely bypassed by projections from the DCN and SOC. The VNLL projects mainly to the ipsilateral central nucleus of the IC (ICC). As it receives its input mainly from the octopus cells, it exhibits temporally precise but complex responses, which do not disclose an obvious function. An unusual feature of the VNLL is that its tonotopic axis is folded in a three-dimensional helicoidal topology, so that its peripheral laminae are mapped to low frequencies of the IC and the central laminae are mapped to high frequencies (Merchán and Berbel, 1996). This structure was compared with the musical pitch helix and associated with auditory sensitivity to periodicity, and hence, was implicated with the basis for harmonicity detection (Langner, 2015; but see Regev et al., 2019).

The function of the INLL is even less clear than the VNLL and DNLL. The INLL primarily projects to the ipsilateral IC and receives its major projections from the ipsilateral MNTB (inhibitory) and the contralateral AVCN and PVCN (excitatory), but with lighter projections from other ipsilateral CN and SOC nuclei (Kelly et al., 2009; Yavuzoglu et al., 2010). It was shown in bats that high-frequency INLL units can be inhibited by low-frequency non-tonotopic sounds—spectral shaping that most likely carries over to the IC and maybe to the auditory cortex (Yavuzoglu et al., 2010).

There is no generally accepted theory for the function of the lower brainstem nuclei, except perhaps for the localization done by the SOC (e.g., Glendenning and Masterton, 1998). The early bifurcation in the auditory pathways at the CN suggests that its different subnuclei have different functions in processing the stimulus. Studies in the avian analogs of the CN, which are simpler than the mammalian CN²², suggest a simpler role division in the barn owl (Sullivan and Konishi, 1984) and chicken (Warchol and Dallos, 1990). The **cochlear magnocellularis** (the avian homolog to the AVCN) is particularly sensitive to temporal changes, judging by its enhanced phase locking to stimuli. At the same time, it is relatively insensitive to intensity changes and has a small dynamic range. The opposite is the case in the **cochlear angularis** (the avian homolog to the PVCN and the DCN), which is sensitive to intensity changes, but shows negligible phase locking. Somewhat similar role division in the brainstem was hypothesized to be separately processing envelope and temporal fine-structure cues that may underlie the “where” and “what” streams, respectively (Smith et al., 2002). Such streams would be analogous to intensity and phase cues from the avian brainstem.

²²The avian auditory brainstem pathway splits to two well-separated streams, instead of three. In the barn owl, this separation is retained also in the processing of interaural time and level differences until they converge at the inferior colliculus (Takahashi et al., 1984).

However, this dual stream interpretation has been challenged (Zeng et al., 2004), and the separation to these two domains turned out to be nearly impossible to accomplish in practice (see §6.5), so if this functional framing has merit, it has to be refined.

A recent review of subcortical auditory processes by Felix II et al. (2018) makes a case for the hierarchical processing scheme mentioned in §2.3.3. According to their model, several features that are useful in auditory scene analysis—the segregation and grouping of sound streams in complex acoustic environments—emerge in the brainstem and become more salient in pathways downstream. These processes include fundamental frequency extraction and harmonicity detection, gap detection, forward masking, improvement of signal in noise and reverberation, spatial segregation, as well as early selectivity to species-specific vocalizations. Brainstem involvement in speech processing goes counter to the traditional approach that associates it with the cortex alone and attributes only general-purpose processing to subcortical areas (Scott and Johnsrude, 2003). Findings that consider speech specialization in humans generally begin in the IC rather than in the CN nuclei, although they may be constrained by methods that do not allow for recordings in humans further upstream (Krishnan and Gandour, 2009). In general, the brainstem appears to be involved in early extraction of features that are meaningful in the description of sound sources, or acoustic objects, rather than in the extraction of mathematical features per se (Masterton, 1992). It has been noted that the brainstem is where the auditory system stabilizes the signal in face of level fluctuations and noise and generally provides the fidelity that is needed to temporally resolve complex sequences of sound (Young and Oertel, 2018).

2.4.2 Midbrain

The **inferior colliculus** (IC) is the primary auditory nucleus in the midbrain and is considered an “obligatory” pathway, as nearly all major and secondary pathways cross it between the CN and the cortex (Aitkin and Phillips, 1984) (see Figure 2.4). In the IC, the different streams that split at the CN converge while retaining their tonotopy. The topographic layout of the **central nucleus of the IC** (ICC) achieves this with thin layers of neurons that form **isofrequency laminae**, which are sharply tuned (they cover about 0.3 octaves, on average), also at high input levels²³. Each lamina maps to a discrete CF, which is thought to mirror the critical-band concept from psychoacoustics (Schreiner and Langner, 1997; Malmierca et al., 2008). However, there are “fine structure” changes in the frequency tuning around the CF along the extent of the two dimensions of the lamina.

Various cells in the IC were found to have specialized responses that are tuned to features such as the direction of sound, the range of amplitude modulation frequencies, or frequency modulated sweeps (in rats and bats). The two-dimensional shape of the isofrequency laminae suggests that they map additional features of sound that are orthogonal to frequency. One influential observation (in the cat) is that periodicity is one such mapped feature, as cells tuned to different modulation frequency ranges were found to be distributed on the lamina, on an axis orthogonal to frequency (Langner and Schreiner, 1988; Schreiner and Langner, 1988). A spatial map for monaural front-back localization was found in the **external nucleus of the inferior colliculus** (ICX) of the guinea pig and a homologous region in the barn owl’s brain. The ICX and the **dorsal cortex of the inferior colliculus** (ICD) also receive projections from somatosensory and trigeminal inputs and show broadly tuned responses that are not necessarily auditory in function. It has been suggested that in these peripheral nuclei of the IC, the auditory system begins its engagement with predictive coding (stimulus deviance detection), which becomes more extensive in the thalamus and cortex (Carbajal and Malmierca, 2018; Carbajal et al., 2024).

²³Isofrequency laminae are sometimes ascribed to the CN topography as well (Young and Oertel, 2018).

The IC also receives centrifugal efferent projections from the auditory cortex and projects to the CN. These projections can have a localized effect on tuning and sensitivity to specific stimuli. Some efferents from the IC (and the VCN) specifically target the MOC neurons in the ventral nucleus of the trapezoid body (VNTB) and neuromodulate the dynamic range of these cells, which is likely reflected in the MOC effect in the cochlea (Romero and Trussell, 2021).

The unique structure of the IC has led to relatively mature research efforts to pin down its strategic involvement in different hearing functions (Winer and Schreiner, 2005a). A theory about the function of the IC was proposed by Casseday and Covey (1996) (see also Casseday et al., 2002), which was inspired by bat echolocation, but is generalizable to mammals and other vertebrates. The theory has two main hypotheses: “(1) *Tuning processes in the IC are related to the biological importance of sounds.* (2) *The change in timing properties at the IC, from rapid input to slowed output, is related to the timing of specific behavioral responses*” (Casseday and Covey, 1996, p. 312). The authors used five lines of evidence to substantiate these hypotheses: the IC has homologs in all vertebrates, it receives inputs from all auditory brainstem nuclei, it has rife connections to the motor system, it has many neurons that are tuned to highly specific sounds that are behaviorally relevant (like parts of species-specific vocalizations or echolocation, such as frequency sweeps), and there is a change to a much slower neural coding. The role of the slowing down at the IC was suggested to be necessary to match the motor responses (e.g., speech production) and to set a slow pace for the cortex to act upon. Alternatively, it was suggested that the slower processing introduces the necessary delay into the system to allow for adequate processing of sound. Overall, this theory implicates the IC with rather general-purpose roles that are neither completely automatic and low level, nor are they particularly complex or in anyway conscious.

2.4.3 Thalamus and cortex

From the IC, the auditory signal continues to the **medial geniculate body** (MGB) in the thalamus. The MGB receives projections from different modalities and is also implicated in fear response to auditory stimuli due to direct projections to the amygdala. The responses that have been recorded in the MGB are complex and diverse and, somewhat like the IC, are relevant to all auditory signal processing aspects. Although it has traditionally been considered to be an auditory relay layer before the cortex (see §2.3), this view is gradually shifting as it has been revealed that the MGB is tightly integrated with the auditory cortex through multiple diverging and converging (thalamocortical) pathways of great diversity in terms of their functions and connections (Winer, 2011a). Aside from providing the main link between the IC and the cortex, information that passes through the MGB can be shaped and optimized using feedback loops from the cortex (corticothalamic projections). Additionally, after the auditory signal convergence in the IC, the MGB diverges, so that several parallel streams project to different places in the auditory cortex. For example, the MGB features significant convergence of receptive fields²⁴ that are more acoustically complex than in the brainstem and IC and culminate in spectrotemporal maps in the auditory cortex (Miller et al., 2001).

The **auditory cortex** is found in the upper surface of the temporal lobe and is subdivided to **core**, **belt**, and **parabelt** areas that are also called the primary, secondary, and tertiary auditory cortex (A1, A2, and A3), respectively. The core projects to the obligatory belt area. The parabelt is connected to the frontal lobe and eye movement control areas. All three areas receive inputs from different areas of the MGB and from one another. Tonotopy occurs independently in several fields

²⁴The **receptive field** refers to the range of input parameters that triggers a response in a neuron. For example, in a central auditory neuron, this may relate to the range of frequencies and interaural timing differences. In general, the more central the neuron is, its receptive field tends to be more specialized to a particular combination of input parameters.

of the core and belt where iso-frequencies are organized in strips. Yet other peripheral areas do not show tonotopy, or rather, they exhibit fragmented frequency maps. This may be indicative of parallel processing that is going on in the auditory cortex. Tonotopic areas have centrifugal projections to other tonotopic areas in the MGB, which potentially gives rise to feedback loops. Other features that have corresponding organized areas in A1 include spatial distribution of monaural or binaural stimuli, areas that show narrow or broad tuning, and cell groups that respond to frequency modulation. Single units also vary in tuning characteristics (narrow or broad), may have multiple peaks, and be influenced by inhibitory or excitatory inputs from other areas. Cortical cells sometimes respond to very specific types of stimuli of growing complexity in areas farther away from A1. Both A1 and A2 also contain cells that are tuned to simultaneous temporal and spectral (**spectrotemporal**) envelope modulations of low frequencies ([Schönwiesner and Zatorre, 2009](#)), which may first appear at the level of the IC, at least in some animals (e.g., [Poon and Yu, 2000](#); [Qiu et al., 2003](#)). A1 cells are shown to specifically respond and adapt to stimuli of different time scales of many orders of magnitude ($10^{-1} - 10^2$ s)—something which does not arise at the processing level of the MGB ([Ulanovsky et al., 2004](#)).

In general, compared to the core areas, the belt seems to have more complex processing that may be related to meaningful signals, such as communication, and is likely engaged in parallel processing (i.e., bypassing A1; cf. [Wang et al., 2024](#)) of complex signals like speech, along with other cortical areas ([Hamilton et al., 2021](#); [Whalen, 2024](#)). Although the auditory system is quite capable without an auditory cortex, it is required for sound localization, sound detection and frequency discrimination tasks—sometimes in a manner that depends on the input to one ear only. Being part of the cortex, the auditory areas seem to exhibit a considerable degree of plasticity, which is predicated on the animal's individual developmental experience and conditions found related to other sensory pathways. Information from the auditory cortex may be used for further decision making and action.

Two complementary approaches for the role of the auditory cortex are that it is either the culmination of hierarchical processing in preceding auditory pathways, or that it serves as a control center that modulates and controls the lower-level input processing via the descending efferent network ([Cariani and Micheyl, 2012](#)). The auditory cortex is hypothesized to be where auditory objects are formed and localized in space and are brought closer to conscious awareness, which itself may be distributed and not confined to one area within the auditory cortex ([Dykstra et al., 2017](#)).

Several hypotheses for the lateralization of the human cortex have been made over the years. For example, that the left hemisphere excels in fast temporal processing ([Schwartz and Tallal, 1980](#)), that the left hemisphere of the auditory cortical areas specializes in temporal processing such as speech, whereas the right hemisphere excels in spectral processing, such as music ([Zatorre et al., 2002](#)), that the left hemisphere specializes in temporal modulations and the right hemisphere in spectral modulation processing ([Flinker et al., 2019](#)), and that the processing of the left auditory cortex is critical for communication (both language and vocalizations; [Ruthig and Schönwiesner, 2022](#)). However, the auditory cortex response has not lent itself to straightforward interpretation as the visual cortex, due to highly specialized and plastic effects, as well as complex subcortical preprocessed input, which is dynamically controlled by the descending centrifugal network ([King and Nelken, 2009](#)).

Although the concept of auditory objects is of prime theoretical interest in this work (§1.4), we will not be dealing directly with the auditory thalamic or the cortical areas. We rather dwell on the concepts of the acoustic object and its corresponding auditory image at the level of the IC, which will likely have some implications on the ideas behind auditory objects.

Formulating a theoretical account of the functions of the auditory cortex is the intersection of a complete hearing theory and a theory of the sensory cortex in general. These theories are in flux and are nowhere near consensual within the research community. A complete review beyond the

above sections and the relevant sections in §1 is outside the scope of this treatise. See [Cariani and Micheyl \(2012\)](#) and [Heilbron and Chait \(2018\)](#) for relevant models and further literature.

2.5 Hearing in humans and other mammals

An overarching goal in much of the auditory research is to better understand human hearing. Detailed knowledge about the biology that underlies the hearing system owes much to various animal species, whose auditory systems are remarkably similar to that of humans. In the case of mammals, there is a great similarity between animals, but also differences in morphological details, which may have effects on physiology that are not always well understood. The most popular species used as animal models are cats, followed (in no particular order) by mice, rats, Mongolian gerbils, chinchillas, guinea pigs, squirrel monkeys (New World monkeys), rhesus macaque monkeys (Old World monkeys), ferrets, marmosets, rabbits, and bats. The latter are often studied as an entirely different specialty within bioacoustics and hearing. The auditory systems of other vertebrates also have significant similarities to humans, so they have been occasionally studied for the same purpose with focus on birds, lizards, and frogs. In this work, notable bird data are brought from the European starling, barn owl, budgerigar, and chicken, as well as bullfrog, tree frog, and bobtail skink lizard data. However, it is useful to remember that the earliest ancestors to the mammalia lineage split from the amniotes (the egg-laying tetrapods that adapted to terrestrial life) before the other vertebrate taxa did—like birds, lizards, or amphibians—so their auditory evolution has been independent of one another throughout the last 320 million years ([Manley, 2017](#)). See [Figure 2.5](#) for a coarse time-line of the ear evolution in amniotes.

Invasive research on the living human hearing organ is out of the question, except for very rare cases in which other medical procedures call for surgical intervention in the ear's area. Thus, invasive experimentation on humans is restricted mainly to studies of cadavers, which allow for anatomical examinations, but only limited physiological and no behavioral observations, to be collected. All other data about human hearing are gathered using indirect methods that may be either objective (e.g., electrophysiology, brain imaging), or psychophysical, which is the only method that can benefit from verbal accounts of the perceptual experience by the listeners. Ideally, invasive and noninvasive methods should all converge to the same results as predicted by theory, should one exist.

On the whole, the auditory systems of all mammals are qualitatively similar but they quantitatively differ: in dimensions, geometries, and sensitivities. However, there are neither extra nor missing auditory organs in any known mammalian order. The differences tend to be striking in the external ears and become more nuanced in the central pathways. One interesting difference is between hearing generalists and hearing specialists, which appears to relate to their cochlear structure—its geometry and organ of Corti configuration. The main focus of the review below is on the differences in auditory systems between mammals, whereas commonalities should be taken from the general description of the hearing organs in §2.2 and §2.4.

The subsection about the external and middle ears is heavily based on [Rosowski \(1994\)](#) and that about the inner ear is based on [Echteler et al. \(1994\)](#), in the same volume.

2.5.1 Outer and middle ear differences

The most prominent auditory differences between mammalian species are visible in their peripheral ears, which to a large extent determine their audible frequency range. Generally, the dimensions, shapes, and structural complexity of all ear parts vary between mammals, often to a great extent. This applies to the shapes of the pinna and the concha, the ear canal length, the diameter (cross section), the canal bends, the eardrum shape, the middle ear air cavity, and the ossicles. The

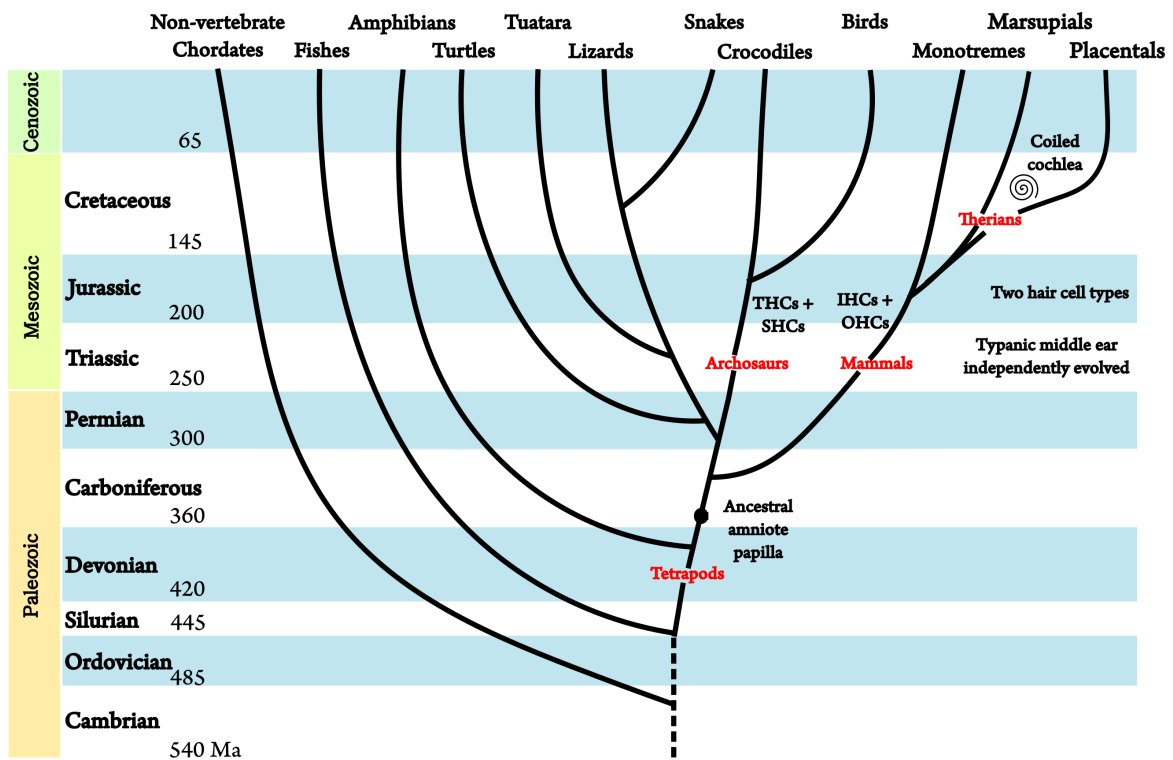


Figure 2.5: A coarse-grained evolutionary tree of the main extant animal clades. The ancestral amniote ear contained a papilla that later evolved into the cochlea in mammals. The tympanic middle ear evolved independently in different animal classes. Similarly, a specialization of hair cells to two types also evolved independently. In birds, these are **tall hair cells** (THCs) and **short hair cells** (SHCs). The figures are adapted from [Manley \(2017\)](#) and [Manley and Köppl \(1998\)](#). The additional split in mammals is based on [Luo et al. \(2011\)](#). The numbers represent the estimated start dates of the different geological periods in million years (megaannum, Ma).

maximum audible frequencies are inversely proportional to the body and head sizes of the animal. As a rule of thumb, small and stiff ears are most effective for high frequencies, whereas large and compliant ears are better for low-frequency hearing. Where a species is able to hear both low and high frequencies well (e.g., cats, gerbils), it is indicative that its middle ear had specialized by evolving its geometry to overcome the conductive limitations imposed by its size. The degree of coupling and type of movement between the eardrum and the ossicles also vary between species. The eardrum and stapes footplate size are approximately scaled like the fourth root of the entire animal body mass, and are linearly correlated with one another. Finally, the effectiveness of the acoustic reflex muscles varies—it is most effective in humans and cats below 1 kHz, and in species with stiff middle ear like bats it is effective over a much broader bandwidth, up to 80 kHz. In several mammalian orders either one of the middle-ear muscles was either lost or degenerate (Mason, 2013). For example, the stapedius is the only active reflex muscle in the guinea pig (and perhaps in the chinchilla), but appears to be degenerate and have little-to-no effect. Several subterranean species have independently lost their tensor tympani muscles, although the advantage in that remains unclear.

The morphological variations have a significant effect on the impedance matching and hence on the power transmission efficiency between the external, middle, and inner ears. External ears collect acoustic power in a frequency-dependent manner, whose efficiency peaks between 10% and 100% somewhere above 1.5 kHz, depending on the particular mismatch between the external ear radiation impedance and the middle ear input impedance of the animal. The impedance matching between the outer and the middle ears is generally poor below 1 kHz and its most effective frequency is usually around 2 kHz, depending on the species. The larger is the area of the eardrum and the air cavity in the middle ear, the smaller is the total middle ear stiffness that dominates its impedance. However, the interdependence of the various parts of the middle ear may be too complex to follow simple laws and the individual anatomy of the ear may be required to determine its sound conduction properties. So, in some cases the stiffness of the eardrum dominates, whereas in other cases the air cavity stiffness dominates. The total transfer of power from the diffuse acoustic field to the oval window, which has the cochlear impedance as its load, is far from ideal and it peaks only at high frequencies, depending on the species. For example, for the cat it is highest at 2–10 kHz and for humans it is much lower and peaks at 1–4 kHz. In the case of gerbils, they excel in low-frequency (<1 kHz) sound collection (which tends to be poor for most animals) that is attributed to their hypertrophied eardrum and middle ear cavity.

The geometry of the pinna and the relative positioning of the ears also affects the directional dependence of incoming sound. For example, the human ears have small pinnae relative to the head size and they are located at its sides, whereas the cat pinnae protrude above it while the ears are still at the sides. Other animals have much larger ear flange sizes relative to their body size, such as the wallaby that has an exceptionally large pinna size to its body. Some mammals can move their external ears and further optimize them for directional localization. In cats, the asymmetrical pinna movement was shown to be coordinated with their eye movements, and the dynamic pinna movement was suggested to improve sound processing (Populin and Yin, 1998). Scattering of low frequencies is affected by the body shape as well, which in the case of human is caused by the torso below about 600–800 Hz. The external ear resonance frequency also varies between species, and is generally higher with shorter ear canals, but depends on incoming sound direction and on additional geometrical features of the ear. In general, large ears collect higher sound power at all frequencies and are more directional at low frequencies than small ears.

Marine mammals evolved independently from terrestrial mammalian orders several times, so specializations for underwater hearing in the different marine taxa do not necessarily have much in common. However, the challenges of hearing in water had to be overcome by all marine mammals, which evolutionarily converged to similar phenotypic solutions, arising from similar genomic expres-

sion (Foote et al., 2015). Theoretically, the aqueous medium has almost identical impedance to the cochlear perilymph, so the impedance matching function may no longer be as necessary underwater as it is overground, but at least in cetaceans the middle ear cavity is filled with air, which allows for pressure gradient separation of the oval and round windows (Ketten, 1998). Thus, the middle ear function is more ambiguous in these animals. Some of them are active terrestrially as well (primarily pinnipeds—seals, sea lions, and walrus), so their ears still have to function both in air and in water, which makes the middle ear necessary. To this effect, pinnipeds are able to regulate the contact with the acoustic medium, as they are equipped with an epithelial layer, which is controlled by vascular activity, that can open and close the ear canal. In general, marine mammals do not have pinnae, apart from sea lions and otters, although some vestigial remnants may be present. In contrast, the ear canal openings of cetaceans is either filled with wax, or is altogether absent. Both cetaceans and pinnipeds have special lining (cavernous mucosa) inside their middle ears to regulate its pressure while diving. Due to the large external pressures, cetaceans evolved massive ossicles in comparison with terrestrial animals. Underwater, it is likely that sound arriving to the inner ear has a dual pathway both through the bones and the tissue. In odontocetes (toothed whales, dolphins, and porpoises), one pathway (through the jaw) may be more suitable for ultrasonic echolocation, whereas the other (through the blocked ear canal) for low-frequency communication (Ketten, 1992; Popov et al., 2008).

2.5.2 Cochlear and auditory nerve differences

Mammals are unique among vertebrates in their high frequency (> 10 kHz) hearing capabilities (but see Nothwang, 2016, for counterexamples). The specific frequency map between the CF to the relative position of the BM can be described using this power law (Greenwood, 1990):

$$f = A(10^{ax} - k) \quad (2.1)$$

with f being the frequency, and A , a , and k are constants specific to the animal. For humans, $A = 165.4$, $k = 0.88$, and $a = 0.06$ when x is in millimeters. This power law is common to many mammals, for which it only differs in the parameters that scale the specific audio bandwidth they respond to according to the BM length. It entails a logarithmic frequency map, as more BM length is dedicated per unit frequency to low frequencies than to high frequencies. So using this formula, it is always possible to express the distance x in relative units to obtain scale-free cochlea that only differs in the audio range. In this case, x would be the proportion distance between 0 and 1, whereas $a = 2.1$ in all animals.

The resonance itself is most conveniently described as a bandpass filtering operation on the incoming sound. The human BM length is taken to be 35 mm on average, although a recent meta-analysis determined it is 33.088 ± 0.452 mm for the highly variable population average (Atalay et al., 2020). However, studies contributing to this figure may have not taken into account the difference between the BM and the cochlear duct lengths due to the helicotrema, which is 1.6 mm on average (Helpard et al., 2020). Furthermore, it was found that the interindividual morphological differences in the spiral shape of the human cochlea are substantial (Pietsch et al., 2017).

Mammals can be roughly categorized in two groups, in terms of their hearing. **Hearing generalists** are those whose hearing threshold is smooth and they do not concentrate on particular spectral regions. Humans, cats, mice, and guinea pigs belong to this category. In contrast, **hearing specialists** have spectral regions of greater fidelity that stand out with respect to their entire audible spectrum. Hearing specialists include bats, rats, dolphins, and gerbils. These preferential spectral regions appear to be behaviorally motivated at least in some of these species. When they exist, these regions are characterized by a relatively constant stiffness in the BM (achieved by unvarying

width and thickness gradient of the BM), along with a discontinuous geometry where the stiffness changes quickly over the cochlear length. Therefore, in hearing specialists, relatively narrow spectral regions occupy large portions of the cochlear length. In bats, these narrowband regions correspond to the second harmonic of their constant-frequency echolocation, where this region is referred to as the **auditory fovea** and it is tonotopically conserved also in the auditory pathways (Covey, 2005). As a result, cochlear power-law scaling applies much better to the generalists than to the specialists.

The cochlea of different mammalian species can be visually distinguished by the geometry of the bony structure—the number of turns in the spiral (2–4), and in the length, width, and thickness of the basilar membrane. None of these parameters, however, correlates well with the basic hearing variables, such as the audible range, or its limits, if the specialists and generalists are mixed, or if large terrestrial animals are included. If only the generalists are included, then longer BMs tend to correlate with lower high- and low-frequency cutoffs. There are indications, though, that the low-frequency cutoff is determined by the size of the opening of the helicotrema that connects the vestibular and tympanic scalae—the larger it is, the higher the cutoff is. High frequency in ten primates (including humans) was robustly shown to be inversely correlated with cochlear volume, independently of body mass, which is itself highly correlated with the BM length (Kirk and Gosselin-Ildari, 2009). Another detailed morphological and audiological survey of the cochleas of 33 different mammalian (therian) species—primarily rodents and primates—suggests that rodents may have evolved to have extended low-frequency audible range through a distinctive “tower-shaped” cochlea (i.e., more elongated in volume) that is achieved through extra coiling (Del Rio et al., 2023). In contrast, for similar cochlear lengths, primates have extended high-frequency audible range by having wider cochleas.

Another difference between mammals that has been recently highlighted is the area of the cochlear partition that supports the traveling wave. Traditionally, it has been considered to be only the BM, but it turns out that the structures that connect the BM to the cochlear wall—a soft **bridge** that connects the BM to the hard plate-like **osseous spiral lamina** (see Figure 2.3)—also vibrate and may even occupy a larger relative surface vibrating area than the BM (Raufer et al., 2019). In humans, the resonant peak of the transverse movement is located underneath the inner pillar cells—in the interface between the bridge and BM. In other mammals, the relative proportion of the vibrating parts and the location of the bridge and the peak relatively to the hair cells can be dramatically different than in humans (Raufer et al., 2019, Supplementary information). The effect of these differences on cochlear models is unknown, but may be significant²⁵.

In various cetaceans, the BM is supported by the osseous spiral lamina in the basal turn of the cochlea, which (along with other morphological differences) gives rise to ultrasonic hearing range that violates the generalists’ scaling law as well (Ketten, 1992, 1998). The basal stiffness itself seems to be correlated with the high-frequency limit of the animals, whereas the apical stiffness is not correlated with the low-frequency limit. Correlations were observed between the thickness of the TM and the pronounced frequency regions in bats and rats.

The length and the quantity of the hair cell bodies and stereocilia vary between different animals. The length of hair cells of all types increases toward the apex. The variation is larger in OHCs (compared to IHCs), whose maximum length in the apex is correlated with the low-frequency limit of the animal. There are on average 3000–3500 IHCs in humans, 2600 in cats, 960 in rats, and about 1100 in gerbils (Úlehlová et al., 1987; Nadol Jr, 1988; Hutson et al., 2021). Another prominent variation between animals is in the number of OHC rows, which is at least three (as in humans), but can be up to six in some rat species in certain cochlear regions. While there may be 11000–16000 OHCs in total in humans, there are less than 10000 in the cat, about 4600 in the gerbil, 3500 in

²⁵In this work we shall refer to the BM as the substrate of the traveling wave. As we will not be concerned with detailed mechanical modeling, this should be understood more generally, as the BM and any connecting structure that vibrate with the traveling wave.

the rat, and 2400 in the guinea pig (Hutson et al., 2021; Nadol Jr, 1988). The number of cilia in each cell is also widely different and is highest for humans and monkeys (up to 150 in the base and 46 in the apex) and less in rats and cats. In the guinea pig, there is almost no variation in cilia count between the base and apex. Several additional morphological differences were found inside hearing specialists' organ of Corti in the form of hypertrophied supporting cells, different TM shapes, a second spiral lamina in some species, and other finer differences (Echteler et al., 1994, pp. 158–162). It is interesting to note that while all vertebrates appear to have some types of hair cells with bundle motility, electromotility is uniquely found in mammalian OHCs (Peng and Ricci, 2011). Another interesting point that has been recently found is that the OHCs and their bundle motility in mice, and perhaps in other small mammals, is required for hearing ultrasound above 16 kHz (Li et al., 2021).

The innervation of the IHCs is determined by the morphology of the spiral ganglion cells and their synaptic terminals. In humans, each IHC is innervated by about half the number of fibers (9–11) than in the cat (20–26), but each human nerve terminal has multiple synapses (15–16) and only one synapse in the cat (Nadol Jr, 1988). The differences in the numbers are reflected also in the total number of spiral ganglion cells in the cochlea, which is 25000–30000 in humans, 45000–58000 in cats, and 15800 in rats. Furthermore, the number of fibers is about 31000 in humans and rhesus monkeys, 52000 in cats, but only 24000 in guinea pigs. It is probably the highest in dolphins, whose exceptional hearing may be superior to humans, where the bottlenose dolphins have approximately 105000 ganglion cells in their cochlea (Ketten, 1992, Table 35.1). Significant structural differences in the bony compartment that houses the spiral ganglia (Rosenthal's canal) of echolocating bats of the suborder Yangochiroptera have enabled larger and more numerous ganglia to evolve, which may have been key in the diversity of echolocation displays that are found in this suborder compared to Yinpterochiroptera (Sulser et al., 2022). The last factoid we shall mention is that only 5% of the cell bodies in humans are myelinated, whereas in cats it is about 95% (Nadol Jr, 1988, Table VI). See Nayagam et al. (2011) for additional species-specific data.

2.5.3 Central differences

Few studies systematically compared the central auditory pathways between mammals. Because of the relative obscure function of most auditory nuclei, it is not obvious what the most informative level of comparison may be, once we factor out the differences in the various receptive fields of the different cells types that may be learned and determined by the animal's environment. Glendenning and Masterton (1998) compared the absolute and relative sizes of the ten most prominent ascending auditory nuclei in a sample of 53 mammalian species²⁶. This study proposed that, in first approximation, the size is indicative of the relative importance of the particular nuclei for the animal. Perhaps the most striking example is that the MSO appeared absent in both mice and hedgehogs, although at least in the case of mice it probably has to do with the method, since the MSO does exist (Fischl et al., 2016). Due to their small head size, ITD cues are unavailable to mice (and other small mammals), and they have to rely on high-frequency ILD cues for localization. The relative size of other nuclei follows a relatively robust mean, in which the IC is by far the largest nucleus and the MGB is the second largest. In some species this is reversed, as the MGB is slightly larger or similar in size to the IC. It is interesting to note that both in the albino rat and the rhesus monkey, it was

²⁶This set does not constitute an unbiased sample of all mammals. For example, the marsupial class is over-represented compared to placentals, while monotremes are altogether absent. Within the placentals, the orders of rodents and chiroptera (bats) are under-represented, while primates are over-represented. However, this sample covers many of commonly used animal models and reveals unmistakable patterns among them, which are more than satisfactory in the present context.

found that the IC has the highest glucose consumption of all auditory system nuclei, followed by the auditory cortex, and then the MGB (Sokoloff et al., 1977; Kennedy et al., 1978). Most animals have uniform CN size relative to the entire auditory system, with the feathertailed glider (a small marsupial) having an unusually large DCN and bats having a large AVCN—measured relative to the total CN, or total auditory system size. However, the interspecies variations of the DCN, PVCN, and AVCN seem to be large. The AVCN and PVCN division itself is not as well-distinguished in humans as it is in other mammals (Moore and Osen, 1979). The MSO of cats, llamas, and foxes is untypically larger than the LSO, which in most species is larger than the MSO.

Although echolocation has been documented in several animal species that evolutionarily converged to a similar orientation principle—notably, bats and toothed-whales, but also some types of rodents and birds (Shen et al., 2012; Parker et al., 2013; He et al., 2021)—it has been studied most intensively in bats. Echolocating bats have highly specialized hearing that makes use of the same auditory nuclei, but sometimes in a different way than all other mammals (Covey, 2005). Some of their auditory organs are hypertrophied—much larger than would be suggested by their brain size—like the cochlea, the IC, the auditory cortex, the VNLL, and the INLL. Notably, the bat's MSO seems to be used for monaural tasks that require temporal precision, rather than for interaural time difference detection, as is the case in other mammals that can hear low frequencies. In some bat species, the VNLL features cells that respond only to particular patterns of modulated sound, with high temporal precision. The IC in echolocating bats is thought to relate their emitted frequency-modulated pulse to the delayed echo from the environment using specialized **delay-tuned** (or “FM-FM”) neurons (that are also found in the INLL, VNLL, MGB, and A1; Wenstrup and Portfors, 2011). This information is then used to derive the distance to the target that is passed on to the thalamus and to the auditory cortex, where pulse-echo maps are formed (organized by the delay time and the FM harmonic that is analyzed) that can inform decision making. Furthermore, dense projections from the IC to premotor areas (pretectal and pontine nuclei) can rapidly stir motor action in flight and dynamically adjust subsequent vocalizations with respect to the target.

A more detailed comparison of the human and cat brainstem nuclei revealed local morphological differences in dimensions (larger in human) and several underdeveloped auditory nuclei in humans—the LSO, the MNTB, and the VNLL (Moore, 1987). The latter is by far the most poorly developed nucleus in humans, in line with other primates and new-world monkeys, and in opposition to bats and porpoises that have highly-developed VNLL. Nevertheless, a double helical structure of the VNLL was identified in humans too that has 7–8 turns—each turn is thought to correspond to one octave (Langner, 2015, pp. 174–176). The human MNTB is also notoriously difficult to identify, but it was argued to positively exist in Grothe et al. (2010, Appendix A). The implications of these differences and others on the morphogenic and cytoarchitectural levels are unknown.

At the level of the auditory cortex there is great morphological variability with different tonotopic maps abound, whose frequency axes are similar down to a scaling factor, despite different boundary shapes (Goldstein and Knight, 1980; Merzenich and Schreiner, 1992). The exception is, yet again, echolocating bats, whose auditory fovea gives rise to tonotopic maps with magnified areas of the foveal narrowband range. Even mammals with relatively primitive cortex like hedgehogs and possums have an auditory cortex with similar responses to more developed and fully laminated cortices (Gates and Aitkin, 1982; Batzri-Izraeli et al., 1990). More generally, it is probably relevant to note that the corpus callosum, which connects the left and right hemispheres, is found only in placental mammals, but not in marsupials and monotremes, or in other vertebrates (Kaas, 2013)—something that undoubtedly has to have some effect on auditory perception and processing as well.

Finally, the scaling of the auditory system relative to the size of the brain is also compressed, as its absolute size is largest in humans, while its relative size to the brain is smallest. It is the opposite in bats that have the largest auditory system relative to their brain size.