

Part I

Introductory Chapters

COPYRIGHTED MATERIAL

1

Ear and Hearing

1.1 The ear

The human ear has striking abilities of detecting and differentiating sounds. It is sensitive to a wide range of frequencies as well as intensities and has an extremely high temporal resolution (for detailed descriptions see, e.g., Geisler, 1998; Moore, 2008; Pickles, 2008; Plack, 2005; Cook, 2001). The ear consists of three parts: The outer, the middle, and the inner ear. The outer ear acts as a receiver and filters sound waves on their way to the ear drum (tympanic membrane) via the ear canal (meatus), amplifying some sounds and attenuating others (depending on the frequency and direction of these sounds). Sound waves (i.e., alternating compression and rarefaction of air) cause the tympanic membrane to vibrate, and these vibrations are subsequently amplified by the middle ear. The middle ear is composed of three linked bones: The malleus, incus, and stapes. These tiny bones help transmit the vibrations on to the oval window of the cochlea, a small membrane-covered opening in the bony wall of the inner ear and the interface between the air-filled middle-ear and the fluid-filled inner ear (Figure 1.1).

The cochlea has three fluid-filled compartments, the scala tympani, the scala media, and the scala vestibuli (which is continuous with the scala tympani at the helicotrema). Scala media and scala tympani are separated by the basilar membrane (BM). The organ of Corti rests on the BM and contains the auditory sensory receptors that are responsible for transducing the sound stimulus into electrical signals. The vibration of the stapes results in varying pressures on the fluid in the scala vestibuli, causing oscillating movements of scala vestibuli, scala media

4 BRAIN AND MUSIC

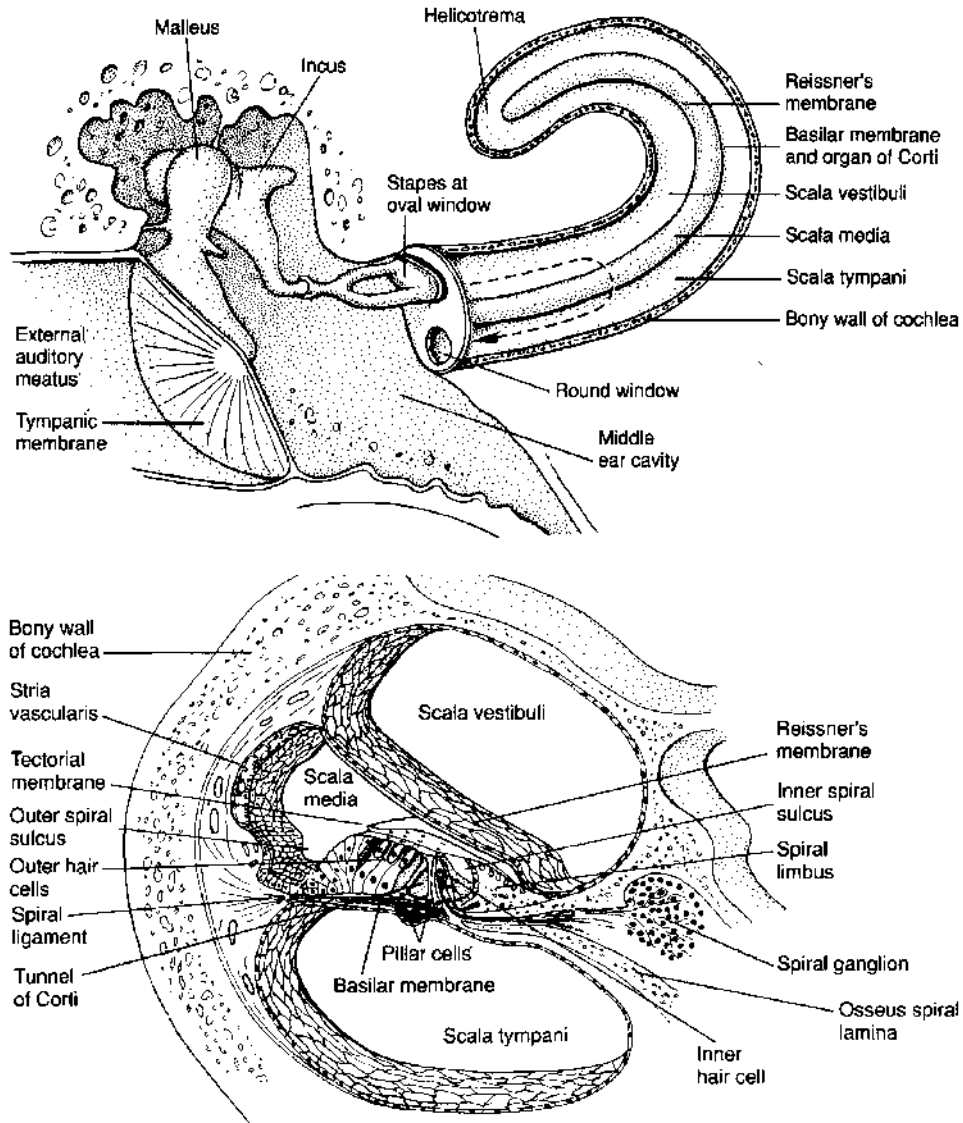


Figure 1.1 Top: The major parts of the human ear. In the Figure, the cochlea has been uncoiled for illustration purposes. Bottom: Anatomy of the cochlea (both figures from Kandel *et al.*, 2000).

(including BM) and scala tympani (for detailed descriptions see, e.g., Geisler, 1998; Pickles, 2008).

The organ of Corti contains the sensory receptor cells of the inner ear, the hair cells (bottom of Figure 1.1). There are two types of hair cells, inner hair cells and outer hair cells. On the apical surface of each hair cell is a bundle of around 100 stereocilia (mechanosensing organelles which respond to fluid motion or fluid

pressure changes). Above the hair cells is the tectorial membrane that attaches to the longest stereocilia of the outer hair cells. The sound-induced movement of the scalae fluid (see above) causes a relative shearing between the tectorial membrane and BM, resulting in a deflection of the stereocilia of both inner and outer hair cells. The deflection of the stereocilia is the adequate stimulus of a hair cell, which then depolarizes (or hyperpolarizes, due to the direction of deflection) by opening an inward current (for detailed information see Steel & Kros, 2001).

The inner hair cells then release glutamate (Nouvian *et al.*, 2006)¹ at their basal ends where the hair cells are connected to the peripheral branches of axons of neurons whose bodies lie in the spiral ganglion. The central axons of these neurons constitute the auditory nerve. The release of glutamate by the hair cells excites the sensory neurons and this in turn initiates action potentials in the cell's central axon in the auditory nerve. Oscillatory changes in the potential of a hair cell thus result in oscillatory release of transmitter and oscillatory firing in the auditory nerve (for details see, e.g., Pickles, 2008; Geisler, 1998). The duration of an acoustic stimulus is encoded by the duration of activation of an auditory nerve fibre.

Different frequencies of sounds are selectively responded to in different regions of the cochlea. Each sound initiates a travelling wave along the length of the cochlea. The mechanical properties of the basilar membrane vary along the length of the cochlea; the BM is stiff and thin at the basal end (and vibrates more to high frequency sounds, similar to the high e-string on a guitar, which resonates at a sound frequency of ~ 330 Hz), whereas at the apex the BM is thicker and less stiff (and resonates at sounds with lower frequencies, similar to the low e-string on a guitar, which resonates at a sound frequency of ~ 82 Hz). Different frequencies of sound produce different travelling waves with peak amplitudes at different points along the BM. Higher frequencies result in peak amplitudes of oscillations of the BM that are located nearer to the base of the cochlea, lower frequencies result in oscillatory peaks near the apex of the cochlea (for more details see, e.g., Pickles, 2008; Geisler, 1998).

The outer hair cells specifically sharpen the peak of the travelling wave at the frequency-characteristic place on the BM (e.g., Fettiplace & Hackney, 2006). Interestingly, outer hair cells achieve the changes in tuning of the local region in the organ of Corti by increasing or decreasing the length of their cell bodies (thereby affecting the mechanical properties of the organ of Corti; Fettiplace & Hackney, 2006). This change in length is an example of the active processes occurring within the organ of Corti while processing sensory information. Moreover, the outer hair cells are innervated by efferent nerve fibres from the central nervous system, and it appears that the changes in length are at least partly influenced by top-down processes (such processes may even originate from neocortical areas of the brain).

¹ The postsynaptic receptors at the afferent synapse to the inner hair cells have been identified as AMPA receptors, and glutamate transporters have been found in nearby supporting cells that dispose of excess glutamate.

6 BRAIN AND MUSIC

Therefore, the dynamics of the cochlea (determining the processing of acoustic information) appears to be strongly influenced by the brain. The dynamic activity of the outer hair cells is a necessary condition for a high frequency-selectivity (which, in turn, is a prerequisite for both music and speech perception).

Corresponding to the tuning of the BM, the frequency-characteristic excitation of inner hair cells gives rise to action potentials in different auditory nerve fibres. Therefore, an auditory nerve fibre is most sensitive to a particular frequency of sound, its so-called *characteristic frequency*. Nevertheless, an individual auditory nerve fibre (which is innervated by several inner hair cells) still responds to a range of frequencies, because a substantial portion of the BM moves in response to a single frequency. The sound pressure level (SPL, for explanation and medical relevance see, e.g., Moore, 2008) is then encoded (1) by the action potential rate of afferent nerve fibres, and (2) by the number of neighbouring afferent nerve fibres that release action potentials (because the number of neurons that release action potentials increases as the intensity of an auditory stimulus increases). The brain decodes the spatio-temporal pattern consisting of the individual firing rates of all activated auditory nerve fibres (each with its characteristic frequency) into information about intensity and frequency of a stimulus (decoding of frequency information is dealt with in more detail further below).

1.2 Auditory brainstem and thalamus

The cochlear nerve enters the central nervous system in the brain stem (cranial nerve VIII).² Within the brain stem, information originating from the hair cells is propagated via both contra- and ipsilateral connections between the nuclei of the central auditory pathway (for a detailed description see Nieuwenhuys *et al.*, 2008). For example, some of the secondary auditory fibres that originate from the ventral cochlear nucleus project to the ipsilateral superior olivary nucleus and to the medial superior olivary nucleus of both sides (both superior olivary nuclei project to the inferior colliculus). Other secondary auditory fibres project to the contralateral nucleus of the trapezoid body (that sends fibres to the ipsilateral superior olivary nucleus; see Figure 1.2). The pattern of contra- and ipsilateral connections is important for the interpretation of interaural differences in phase and intensity for the localization of sounds.

The inferior colliculus (IC) is connected with the medial geniculate body of the thalamus. The cells in the medial geniculate body send most of their axons via the radiatio acustica to the ipsilateral primary auditory cortex (for a detailed description see Nieuwenhuys *et al.*, 2008). However, neurons in the medial division of the medial geniculate body (mMGB) also directly project to the lateral amygdala (LeDoux, 2000); specifically those mMGB neurons receive ascending inputs from

² This can easily be remembered, because both ‘ear’ and ‘eight’ start with an ‘e’.

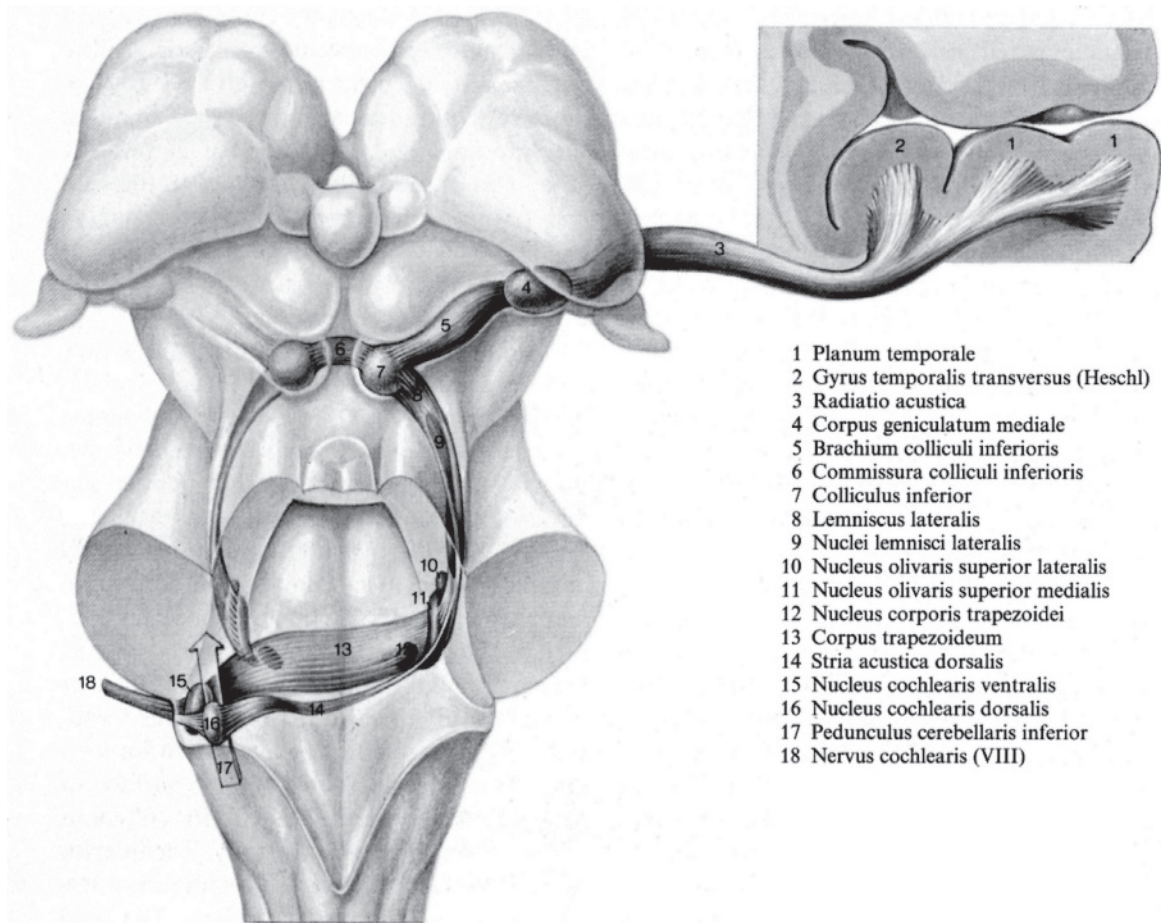


Figure 1.2 Dorsal view of nerve, nuclei, and tracts of the auditory system (from Nieuwenhuys *et al.*, 2008).

the inferior colliculus and are likely to be, at least in part, acoustic relay neurons (LeDoux *et al.*, 1990). The MGB, and presumably the IC as well, are involved in conditioned fear responses to acoustic stimuli. Moreover, already the IC plays a role in the expression of acoustic-motor as well as acoustic-limbic integration (Garcia-Cairasco, 2002), and chemical stimulation of the IC can evoke defence behaviour (Brandão *et al.*, 1988). It is for these reasons that the IC and the MGB are not simply acoustic relay stations, but that these structures are involved in the detection of auditory signals of danger.

What is often neglected in descriptions of the auditory pathway is the important fact that auditory brainstem neurons also project to neurons of the reticular formation. For example, intracellular recording and tracing experiments have shown that giant reticulospinal neurons in the caudal pontine reticular formation (PnC) can be driven at short latencies by acoustic stimuli, most presumably due to

multiple and direct input from the ventral (and dorsal) cochlear nucleus (perhaps even from interstitial neurons of the VIII nerve root) and nuclei in the superior olivary complex (e.g., lateral superior olive, ventral periolivary areas; Koch *et al.*, 1992). These reticular neurons are involved in the generation of motor reflexes (by virtue of projections to spinal motoneurons), and it is conceivable that the projections from the auditory brainstem to neurons of the reticular formation contribute to the vitalizing effects of music, as well as to the (human) drive to move to music (perhaps in interaction with brainstem neurons sensitive for isochronous stimulation).³

1.3 Place and time information

The tonotopic excitation of the basilar membrane (BM),⁴ is maintained as tonotopic structure (also referred to as *tonotopy*) in the auditory nerve, auditory brainstem, thalamus, and the auditory cortex. This tonotopy is an important source of information about the frequencies of tones. However, another important source is the temporal patterning of the action potentials generated by auditory nerve neurons. Up to frequencies of about 4–5 kHz, auditory nerve neurons produce action potentials that occur approximately in phase with the corresponding oscillation of the BM (although the auditory nerve neurons do not necessarily produce an action potential on every cycle of the corresponding BM oscillation). Therefore, up to about 4–5 kHz, the time intervals between action potentials of auditory neurons are approximately integer ratios of the period of a BM oscillation, and the timing of nerve activity codes the frequency of BM oscillation (and thus of the frequency of a tone, or partial of a tone, which elicits this BM oscillation). The brain uses both place information (i.e., information about which part/s, of the BM was/were oscillating) and time information (i.e., information about the frequency/ies of the BM oscillation/s). Note, however, (a) that time information is hardly available at frequencies above about 5 kHz, (b) that place information appears to be not accurate enough to decode differences in frequencies in the range of a few percent (e.g., between a tone of 5000 and 5050 Hz), and (c) that place information alone cannot explain the phenomenon of the pitch perception of tones with *missing fundamentals*⁵ (for details about the *place theory* and *temporal theory* see, e.g., Moore, 2008).

³ A recent study by Zentner & Eerola (2010) suggests that this drive is already present in infants.

⁴ Recall that higher frequencies result in peak amplitudes closer to the base of the cochlea, and lower frequencies in peaks near the apex of the cochlea.

⁵ What is the perceived pitch of a tone consisting, for example, of the frequencies 200 Hz, 300 Hz, 400 Hz and 500 Hz? The answer is 100 Hz (not 200 Hz!), because all partials are integer multiples of a *missing fundamental* frequency of 100 Hz. Therefore, the perceived pitch of a complex tone consisting of the frequencies 400 Hz, 500, 600 Hz, and 700 Hz is also 100 Hz. That is, if a tone has enough overtones, then the fundamental frequency could be filtered out, and the pitch percept would remain unchanged (what would change, however, is the timbre of the sound).

The phenomenon of the perception of a ‘missing fundamental’ is an occurrence of *residue pitch*,⁶ also referred to as *periodicity pitch*, *virtual pitch*, or *low pitch*. The value of a residue pitch equals the periodicity (i.e., the timing) of the waveform resulting from the superposition of sinusoids. Importantly, dichotically presented stimuli also elicit residue perception, arguing for the notion that temporal coding of sounds beyond the cochlea is important for pitch perception. Such temporal coding has been reported for neurons of the inferior colliculus (e.g., Langner *et al.*, 2002) and the auditory cortex (see below);⁷ even neurons in the (dorsal) cochlear nucleus (DCN) are able to represent the periodicity of iterated rippled noise, supporting the notion that already the DCN is involved in the temporal representation of both envelope periodicity and pitch (Neuert *et al.*, 2005). However, note that two (or more) frequencies that can be separated (or ‘resolved’) by the BM, also generate (intermodulation) distortions on the BM with different frequencies, the one most easily audible having a frequency of $f_2 - f_1$ (usually referred to as difference *combination tone*). Usually both mechanisms (BM distortions generating combination tones, and temporal coding) contribute to the perception of residue pitch, although combination tones and residue pitch can also be separated (Schouten *et al.*, 1962).

1.4 Beats, roughness, consonance and dissonance

If two sinusoidal tones (or two partials of two tones with similar frequency) cannot be separated (or ‘resolved’) by the BM, that is, if two frequencies pass through the same *equivalent rectangular bandwidth* (ERB; for details see Moore, 2008; Patterson & Moore, 1986),⁸ then the two frequencies are added together (or ‘merged’) by the BM. This results in an oscillation of the BM with a frequency equal to the mean frequency of the two components, and an additional *beat*⁹ (see also von Helmholtz, 1870). Such beats are regular amplitude fluctuations occurring due to the changing phase relationship between the two initial sinusoids, which results in the phenomenon that the sinusoids alternately reinforce and cancel out each other. The frequency of the beat is equal to the frequency difference between the two initial sinusoids. For example, two sinusoidal tones with frequencies of 1000 and 1004 Hz add up to (and are then perceived as) a tone of 1002 Hz, with four beats occurring each second (similar to when turning a volume knob up and

⁶ The German term for ‘residue pitch’ is ‘Residualton’.

⁷ As a note of caution, (McAlpine *et al.*, 2000) showed that some neural responses representing periodicity information at the level of the inferior colliculus may simply be due to cochlear (intermodulation) distortions.

⁸ Others have used the term *critical band* (Zwicker, 1961; Zwicker & Terhardt, 1980) or *auditory filter* (for a historical account and details see Moore, 2008).

⁹ The German word for beats with relatively low frequency (roughly below about 15–20 Hz) is *Schwebung*.

10 BRAIN AND MUSIC

down four times in one second). When the beats have higher frequencies (above ~ 20 Hz), these beats are perceived as *roughness* (Plomp & Steeneken, 1968; Terhardt, 1974; 1978), and are a sensory basis for the so-called *sensory dissonance* (Terhardt, 1976, 1984; Tramo *et al.*, 2001). Western listeners tend to judge two sinusoidal tones as consonant as soon as their frequency separation exceeds about one ERB (Plomp & Levelt, 1965), which is typically between 11% and 17% of the centre frequency.

Ernst Terhardt (1976; 1984) distinguished two components of musical consonance/dissonance, namely *sensory consonance/dissonance* and *harmony*.¹⁰ According to Terhardt, sensory consonance/dissonance represents the graded absence/presence of annoying factors (such as beats and roughness). Others (Tramo *et al.*, 2001) argued that consonance is also a positive phenomenon (not just a negative phenomenon that depends on the absence of roughness), one reason being that residue pitches produced by the auditory system contribute to the percept of consonance.¹¹ Tramo *et al.* (2001) argue that, in the case of consonant intervals, the most common interspike interval (ISI) distributions of auditory nerve fibres correspond (a) to the F0 frequencies of the tones, as well as (b) to the frequency (or frequencies) of the residue pitch(es). Moreover (c), all or most of the partials can be resolved. By contrast, for dissonant intervals, the most common ISIs in the distribution do not correspond (a) to either of the F0s, nor (b) to harmonically related residue pitch(es). Moreover (c), many partials cannot be resolved.

Harmony, according to Terhardt, represents the fulfilment, or violation, of musical regularities that, given a particular musical style, govern the arrangement of subsequent or simultaneously sounding tones ('tonal affinity, compatibility, and fundamental-note relation', Terhardt, 1984 p. 276).¹² The degree to which *harmony* is perceived as un/pleasant is markedly shaped by cultural experience, due to its relation to music- (and thus presumably also culture-) specific principles.

Sensory dissonance (i.e., the 'vertical dimension of harmony'; Tramo *et al.*, 2001) is universally perceived as less pleasant than consonance, but the degree to which sensory consonance/dissonance is perceived as pleasant/unpleasant is also significantly shaped by cultural experience. This notion has recently received support by a study carried out in Cameroon with individuals of the Mafa people who had presumably never listened to Western music before participating in the experiment (Fritz *et al.*, 2009). The Mafa showed a significant preference for original Western music over continuously dissonant versions of the same pieces.

¹⁰ (Tramo *et al.*, 2001) use the terms *vertical* and *horizontal* dimensions of harmony instead. They restrict use of the terms *consonance* and *dissonance* to the vertical dimension of harmony.

¹¹ Note that merely a critical band account of consonance as the absence of roughness cannot explain why in experiments with pure tones the interval of a tritone is perceived as less consonant (or more dissonant) than the fourth or the fifth, although both pitches can clearly be resolved by the BM (for a review see Tramo *et al.*, 2001).

¹² (Tramo *et al.*, 2001) use the term 'horizontal dimension of harmony' instead.

Notably, the difference in normalized pleasantness ratings between original music and the continuously dissonant versions was moderate, and far smaller than those made by a control group of Western listeners. That is, both Western and Mafa listeners preferred more consonant over continuously dissonant music, but whereas this preference was very strong in Western listeners, it was rather moderate in the Mafas. This indicates that the preference for mainly consonant music over continuously dissonant music is shaped by cultural factors.¹³

Beating sensations can not only occur monaurally (i.e., when different frequencies enter the same ear), but also binaurally (i.e., when each ear receives different frequencies, for example one frequency entering one ear, and another frequency entering the other ear). Binaural beats presumably emerge mainly from neural processes in the auditory brainstem (Kuwada *et al.*, 1979; McAlpine *et al.*, 2000), which are due to the continuously changing interaural phase that results from the superposition of two sinusoids, possibly related to sound localization.¹⁴ Perceptually, binaural beats are somewhat similar to monaural beats, but not as distinct as monaural beats. Moreover, in contrast to monaural beats (which can be observed over the entire audible frequency range) binaural beats are heard most distinctly for frequencies between 300 and 600 Hz (and they become progressively more difficult to hear at higher frequencies; for details see Moore, 2008).

1.5 Acoustical equivalency of timbre and phoneme

With regard to a comparison between music and speech, it is worth mentioning that, in terms of acoustics, there is no difference between a phoneme and the timbre of a musical sound (and it is only a matter of convention if phoneticians use terms such as ‘vowel quality’ or ‘vowel colour’, instead of ‘timbre’).¹⁵ Both are characterized by the two physical correlates of timbre: *Spectrum envelope* (i.e.,

¹³ Interestingly, the cultural influence on preference of consonance/dissonance works both ways: Individuals who listen a lot to music with high degree of dissonance begin to prefer higher degrees of dissonance in music. This is reminiscent, for example, of the un/pleasantness caused by capsaicin (the alkaloid that makes paprika and chili taste hot); capsaicin is universally perceived as less pleasant than sugar (Rozin & Schiller, 1980), but individuals develop strong, culture-specific preferences for strong spices. In fact, adults across the world daily ingest substances that are innately rejected, such as bitter substances, or substances irritating the oral mucosa (e.g., coffee, beer, spirits, tobacco, and chili pepper; Rozin & Schiller, 1980).

¹⁴ Contrary to what sellers of so-called *i-dosing* audio-files promise, it is almost certain that binaural beats themselves cannot evoke brain states that are even remotely comparable with those induced by drugs such as heroin, marijuana, etc. There is also lack of scientific evidence indicating that binaural beats have any systematic effect on relaxation, anxiety-reduction, etc.

¹⁵ When two sounds are perceived as having the same pitch, loudness, duration, and location of origin, and ‘a difference can still be heard between the two sounds, that difference is called timbre’ (e.g., Moore, 2008). For example: Imagine that a clarinet, a saxophone, and a piano successively play a middle C at the same location, with the same loudness and the same duration. Each of these instruments has a unique sound quality. This difference is called timbre, tone colour, or simply sound

differences in the relative amplitudes of the individual harmonics) and *amplitude envelope* (also referred to as amplitude contour or energy contour of the sound wave, i.e., the way that the loudness of a sound changes, particularly with regard to the attack and the decay of a sound).¹⁶ Aperiodic sounds can also differ in spectrum envelope (see, e.g., the difference between /f/ and /s/), and timbre differences related to amplitude envelope play a role in speech, e.g., in the shape of the attack for /b/ vs. /w/ and /f/ vs. /tʃ/.

1.6 Auditory cortex

The *primary auditory cortex* corresponds to the transverse gyrus of Heschl (or gyrus temporalis transversus) which is part of the superior temporal gyrus (STG). Most researchers agree that the primary auditory cortex (corresponding to Brodmann's area 41) consists of three sub-areas, referred to as AI (or A1), R, and RT by some authors (e.g., Kaas & Hackett, 2000; Petkov *et al.*, 2006; see also Figure 1.3), or Te1.0, Te1.1, and Te1.2 by others (Morosan *et al.*, 2001, 2005). The primary auditory cortex (or 'auditory core region') is surrounded by auditory belt and parabelt regions that constitute the *auditory association cortex* (Kaas & Hackett, 2000; Petkov *et al.*, 2006).^{17,18}

Figure 1.3 shows these regions and their connectivity according to the nomenclature introduced by Kaas & Hackett (2000).¹⁹ Note that, unlike what is shown in Figure 1.3, Nieuwenhuys *et al.* (2008) stated that the parabelt region also covers parts of the temporal operculum, that is, part of the medial (and not only the lateral) surface of the STG (p. 613). Nieuwenhuys *et al.* (2008) also noted that the precise borders of the posterior parabelt region (which grades in the left hemisphere into Wernicke's area) are not known, but that 'it is generally assumed that it includes the posterior portions of the planum temporale and superior

quality. There are also many examples of timbre differences in speech. For example, two vowels spoken with the same loudness and same pitch differ from one another in timbre.

¹⁶ E.g., sudden or slow attack or decay, such as in the sounds of plucked vs. bowed stringed instruments. Additional features include microtemporal variations such as jitter (microvariations in the F0 frequency) and shimmer (microvariations in the glottal pulse amplitude), which are also characteristic for both 'phonemes' and 'timbres.'

¹⁷ In terms of Brodmann's nomenclature, the auditory core region appears to correspond to Brodmann's area (BA) 41, the lateral auditory belt region to BA 42, medial belt region to BA 52, and auditory parabelt region to much of BA 22 (Hackett & Kaas, 2004, although parts of BA 22 may also constitute the auditory belt region).

¹⁸ Galaburda & Sanides (1980) reported that (in humans) regions of caudo-dorsal parakoniocortex (PaA c/d) extended from the posterior temporal plane (caudomedial end of the Sylvian fissure), *around* the retroinsular region and then dorsally onto the medial aspect of the parietal operculum. Thus, according to Galaburda & Sanides (1980), auditory cortex can also be found in the parietal operculum.

¹⁹ Others (e.g., Morosan *et al.*, 2001, 2005) refer to these regions as areas Te2.1, Te2.2, Te3, and Te4.

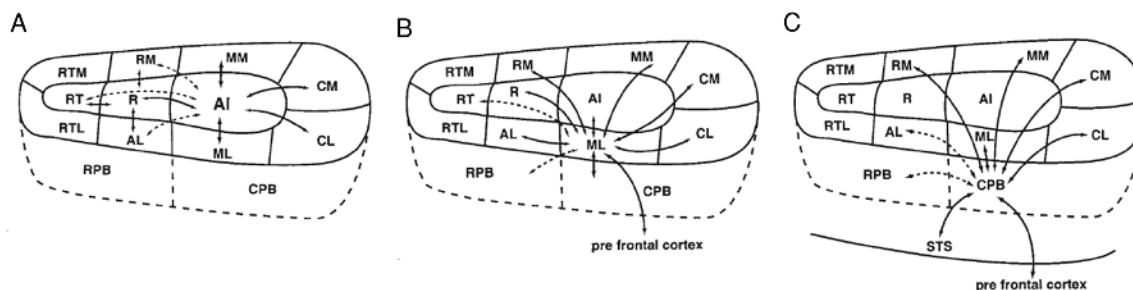


Figure 1.3 Subdivisions and connectivity of the auditory cortex. **(A)** The auditory core region (also referred to as primary auditory cortex) is comprised of the auditory area I (AI), a rostral area (R), and a rostrottemporal area (RT). Area AI, as well as the other two core areas, has dense reciprocal connections with adjacent areas of the core and belt (left panel, solid lines with arrows). Connections with nonadjacent areas are less dense (left panel, dashed lines with arrows). The core has few, if any, connections with the parabelt or more distant cortex. **(B)** shows auditory cortical connections of the middle lateral auditory belt area (ML). Area ML, as well as other belt areas, have dense connections with adjacent areas of the core, belt, and parabelt (middle panel, solid lines with arrows). Connections with nonadjacent areas tend to be less dense (middle panel, dashed lines with arrows). The belt areas also have topographically organized connections with functionally distinct areas in the prefrontal cortex. **(C)** Laterally adjacent to the auditory belt is a rostral (RPB) and a caudal parabelt area (CPB). Both these parabelt areas have dense connections with adjacent areas of the belt and RM in the medial belt (illustrated for CPB by the solid lines with arrows). Connections to other auditory areas tend to be less dense (dashed lines with arrows). The parabelt areas have few, if any, connections with the core areas. The parabelt also has connections with the polysensory areas in the superior temporal sulcus (STS) and with functionally distinct areas in prefrontal cortex. Further abbreviations: CL, caudolateral area; CM, caudomedial area; ML, middle lateral area; RM, rostromedial area; AL, anterolateral area; RTL, lateral rostrottemporal area; RTM, medial rostrottemporal area. Reprinted with permission from Kaas & Hackett (2000).

temporal gyrus, and the most basal parts of the angular and supramarginal gyri? (p. 613–614).

All of the core areas, and most of the belt areas, show a tonotopic structure, which is clearest in AI. The tonotopic structure of R seems weaker than that of AI, but stronger than that of RT. The majority of belt areas appear to show a tonotopic structure comparable to that of R and RT (Petkov *et al.*, 2006, reported that, in the macaque monkey, RTM and CL have only a weak, and RTL and RM no clear tonotopic structure).

The primary auditory cortex (PAC) is thought to be involved in several auditory processes. (1) The *analysis of acoustic features* (such as frequency, intensity, and timbral features). Compared to the brainstem, the auditory cortex is capable of performing such analysis with considerably higher resolution (perhaps with the exception of the localization of sound sources). Tramo *et al.* (2002) reported that a patient with bilateral lesions of the PAC (a) had normal detection thresholds for sounds (i.e., the patient could say whether there was a tone or not), but (b) had elevated thresholds for determining whether two tones have the same pitch or not (i.e., the patient had difficulties detecting minute frequency differences

between two subsequent tones). (2) *Auditory sensory memory* (also referred to as ‘echoic memory’). The auditory sensory memory is a short-term buffer that stores auditory information for a few instances (up to several seconds). (3) *Extraction of inter-sound relationships*. The study by Tramo *et al.* (2002) also reported that the patient with PAC lesions had markedly increased thresholds for determining the pitch direction (i.e., the patient had great difficulties in saying whether the second tone was higher or lower in pitch than the first tone, *even though* he could tell that both tones differed (see also Johnsrude *et al.*, 2000; Zatorre, 2001, for similar results obtained from patients with right PAC lesions). (4) *Stream segregation*, including discrimination and organization of sounds as well as of sound patterns (see also Fishman *et al.*, 2001). (5) *Automatic change detection*. Auditory sensory memory representations also serve the detection of changes in regularities inherent in the acoustic input. Such detection is thought to be reflected electrically as the mismatch negativity (MMN; see Chapter 5), and several studies indicate that the PAC is involved in the generation of the MMN (for an MEG-study localizing the MMN generators in the PAC see Maess *et al.*, 2007). (6) *Multisensory integration* (Hackett & Kaas, 2004), particularly integration of auditory and visual information. (7) *The transformation of acoustic features into auditory percepts*, that is, transformation of acoustic features such as frequency, intensity etc. into auditory percepts such as pitch height, pitch chroma, and loudness.²⁰ It appears that patients with (right) PAC lesions have lost the ability to perceive residue pitch (Zatorre, 1988), consistent with animal studies showing that bilateral lesions of the auditory cortex (in the cat) impair the discrimination of changes in the pitch of a missing fundamental (but not changes in frequency alone; Whitfield, 1980). Moreover, neurons in the anterolateral region of the PAC show responses to a missing fundamental frequency (Bendor & Wang, 2005, data were obtained from marmoset monkeys), and magnetoencephalographic data suggest that response properties in the PAC depend on whether or not a missing fundamental of a complex tone is perceived (Patel & Balaban, 2001, data were obtained from humans). In that study (Patel & Balaban, 2001) phase changes of the auditory steady-state response (aSSR) were related to the pitch percept of a sound.²¹

As mentioned above, combination tones emerge already in the cochlea (generated by the nonlinear mechanics of the basilar membrane), and the periodicity of complex tones is coded in the spike pattern of auditory brainstem neurons.²² That is, different mechanisms contribute to the perception of residue pitch on at least

²⁰ For example, a sound with the frequencies 200 Hz, 300 Hz, and 400 Hz is transformed into the pitch percept of 100 Hz.

²¹ The aSSR is an ongoing oscillatory brain signal resulting from continuous amplitude modulation (AM) of an acoustic stimulus; for example, in the study by Patel & Balaban (2001), complex tones were amplitude-modulated at a rate of 41.5 Hz. The aSSR presumably originates from the PAC (e.g., Ross *et al.*, 2000).

²² Responses in the PAC related to the perception of missing fundamental frequencies in the studies by Bendor & Wang (2005) and Patel & Balaban (2001) are presumably partly due to the periodicity information about the missing fundamental frequency coded in the spike pattern of collicular neurons.

three different levels: (1) On the basilar membrane (BM), (2) in the brainstem (due to temporal coding that leads to a periodicity of the neuronal spike pattern), and (3) in the auditory cortex.²³ However, the studies by Zatorre (2001) and Whitfield (1980) suggest that the auditory cortex plays a more prominent role for the transformation of acoustic features into auditory percepts than the brainstem (or the basilar membrane).

It is also worth noting that neurons in AI are responsive to both sinusoidal ('pure') tones and complex tones, as well as to noise stimuli, whereas areas outside AI become increasingly unresponsive to pure tones, and respond more strongly (or exclusively) to complex tones and noises. Therefore, it seems most plausible that accurate acoustic feature analysis, sound discrimination and pattern organization, as well as transformation of acoustic features into percepts are the results of close interactions between auditory core and belt areas. In addition, the auditory association cortex fulfils a large array of functions (many of which have just begun to be investigated systematically with neuroscientific methods) such as auditory scene analysis and stream segregation (De Sanctis *et al.*, 2008; Gutschalk *et al.*, 2007; Snyder & Alain, 2007), auditory memory (Näätänen *et al.*, 2010; Schonwiesner *et al.*, 2007), phoneme perception (Obleser & Eisner, 2009), voice perception (Belin *et al.*, 2004), speaker identification (von Kriegstein *et al.*, 2005), perception of the size of a speaker or an instrument (von Kriegstein *et al.*, 2007), audio-motor transformation (Warren *et al.*, 2005; Rauschecker & Scott, 2009), syntax processing (Friederici, 2009), or storage and activation of lexical representations (Lau *et al.*, 2008).

With regard to functional differences between the left and the right PAC, as well as neighbouring auditory association cortex, several studies indicate that the left auditory cortex (AC) has a higher resolution of temporal information than the right AC, and that the right AC has a higher spectral resolution than the left AC (Zatorre *et al.*, 2002; Hyde *et al.*, 2008). Furthermore, with regard to pitch perception, Warren *et al.* (2003) report that changes in pitch height as well as changes in pitch chroma (see p. 20 for description of the term 'pitch chroma') activate PAC, but that chroma changes involve auditory belt areas anterior of the PAC (covering parts of the planum polare) more strongly than changes in pitch height. Conversely, changes in pitch height activated auditory belt areas posterior of the PAC (covering parts of the planum temporale) more strongly than changes in pitch chroma.

With regard to the perception of the pitches of melodies, it appears that the analysis of the contour of a melody (which is part of the auditory Gestalt formation)²⁴ particularly relies on the right superior temporal gyrus (posterior rather

²³ But note also that combination tones and residue pitch can be separated (Schouten *et al.*, 1962).

²⁴ The formation of auditory Gestalten follows so-called Gestalt-principles, such as the principle of similarity, of proximity, or of continuation. For example, (1) the single tones of a chord are perceived as one auditory Gestalt (a chord) because they are played at the same time (principle of contiguity); (2) when a melody is played in a high register which is accompanied by chords in a low register, the

16 BRAIN AND MUSIC

than anterior STG), whereas the use of more detailed interval information appears to involve both posterior and anterior areas of the supratemporal cortex bilaterally (Peretz & Zatorre, 2005; Liegeois-Chauvel *et al.*, 1998; Patterson *et al.*, 2002). The planum temporale especially has been implicated in the processing of pitch intervals and sound sequences (Patterson *et al.*, 2002; Zatorre *et al.*, 1994; Koelsch *et al.*, 2009), consistent with the notion that this region is a crucial structure for auditory scene analysis and stream segregation. An introduction to subjective measures of pitch perception is provided in Chapter 3.

tones of the melody are perceived as one Gestalt, and the tones of the chords as another, even if they have the same onsets (principle of proximity); (3) if the same melody is played in a low register by a cello, and the chords are played in a low register on a piano, then the cello tones are perceived as one Gestalt and the chords as another (principle of similarity); (4) if two cellos play two melodies, and both melodies cross, then one melody will be perceived as ascending and the other as descending (principle of continuity).