

stefan koelsch

# brain & music





# Brain and Music



# Brain and Music

Stefan Koelsch

 **WILEY-BLACKWELL**

A John Wiley & Sons, Ltd., Publication

This edition first published 2013  
© 2013 by John Wiley & Sons, Ltd

Wiley-Blackwell is an imprint of John Wiley & Sons, formed by the merger of Wiley's global Scientific, Technical and Medical business with Blackwell Publishing.

*Registered Office*

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

*Editorial Offices*

9600 Garsington Road, Oxford, OX4 2DQ, UK  
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK  
111 River Street, Hoboken, NJ 07030-5774, USA

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at [www.wiley.com/wiley-blackwell](http://www.wiley.com/wiley-blackwell).

The right of the author to be identified as the author of this work has been asserted in accordance with the UK Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book. This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

*Library of Congress Cataloguing-in-Publication Data applied for.*

ISBN hardback: 9780470683408  
ISBN paperback: 9780470683392  
ISBN ePDF: 9781118374023  
ISBN ePub: 9781119943112  
ISBN mobi: 9781119943129

A catalogue record for this book is available from the British Library.

Set in 10.5/13pt Galliard by Aptara Inc., New Delhi, India

First 2013

# Contents

Preface	ix
<b>Part I Introductory Chapters</b>	<b>1</b>
1 Ear and Hearing	3
1.1 The ear	3
1.2 Auditory brainstem and thalamus	6
1.3 Place and time information	8
1.4 Beats, roughness, consonance and dissonance	9
1.5 Acoustical equivalency of timbre and phoneme	11
1.6 Auditory cortex	12
2 Music-theoretical Background	17
2.1 How major keys are related	17
2.2 The basic in-key functions in major	20
2.3 Chord inversions and Neapolitan sixth chords	21
2.4 Secondary dominants and double dominants	21
3 Perception of Pitch and Harmony	23
3.1 Context-dependent representation of pitch	23
3.2 The representation of key-relatedness	26
3.3 The developing and changing sense of key	29
3.4 The representation of chord-functions	30
3.5 Hierarchy of harmonic stability	31
3.6 Musical expectancies	35
3.7 Chord sequence paradigms	36

4	From Electric Brain Activity to ERPs and ERFs	40
4.1	Electro-encephalography (EEG)	43
4.1.1	The 10–20 system	43
4.1.2	Referencing	45
4.2	Obtaining event-related brain potentials (ERPs)	45
4.3	Magnetoencephalography (MEG)	48
4.3.1	Forward solution and inverse problem	49
4.3.2	Comparison between MEG and EEG	49
5	ERP Components	51
5.1	Auditory P1, N1, P2	51
5.2	Frequency-following response (FFR)	53
5.3	Mismatch negativity	54
5.3.1	MMN in neonates	57
5.3.2	MMN and music	57
5.4	N2b and P300	59
5.5	ERP-correlates of language processing	59
5.5.1	Semantic processes: N400	60
5.5.2	Syntactic processes: (E)LAN and P600	63
5.5.3	Prosodic processes: Closure Positive Shift	67
6	A Brief Historical Account of ERP Studies of Music Processing	70
6.1	The beginnings: Studies with melodic stimuli	70
6.2	Studies with chords	74
6.3	MMN studies	75
6.4	Processing of musical meaning	76
6.5	Processing of musical phrase boundaries	77
6.6	Music and action	77
7	Functional Neuroimaging Methods: fMRI and PET	79
7.1	Analysis of fMRI data	81
7.2	Sparse temporal sampling in fMRI	84
7.3	Interleaved silent steady state fMRI	85
7.4	‘Activation’ vs. ‘activity change’	85
	<b>Part II Towards a New Theory of Music Psychology</b>	<b>87</b>
8	Music Perception: A Generative Framework	89
9	Musical Syntax	98
9.1	What is musical syntax?	98
9.2	Cognitive processes	102



9.3	The early right anterior negativity (ERAN)	109
9.3.1	The problem of confounding acoustics and possible solutions	113
9.3.2	Effects of task-relevance	120
9.3.3	Polyphonic stimuli	121
9.3.4	Latency of the ERAN	127
9.3.5	Melodies	127
9.3.6	Lateralization of the ERAN	129
9.4	Neuroanatomical correlates	131
9.5	Processing of acoustic vs. music-syntactic irregularities	133
9.6	Interactions between music- and language-syntactic processing	138
9.6.1	The Syntactic Equivalence Hypothesis	145
9.7	Attention and automaticity	147
9.8	Effects of musical training	149
9.9	Development	151
10	Musical Semantics	156
10.1	What is musical semantics?	156
10.2	Extra-musical meaning	158
10.2.1	Iconic musical meaning	158
10.2.2	Indexical musical meaning	159
	Excursion: Decoding of intentions during music listening	161
10.2.3	Symbolic musical meaning	162
10.3	Extra-musical meaning and the N400	163
10.4	Intra-musical meaning	170
	Excursion: Posterior temporal cortex and processing of meaning	166
10.4.1	Intra-musical meaning and the N5	171
10.5	Musicogenic meaning	177
10.5.1	Physical	177
10.5.2	Emotional	179
10.5.3	Personal	180
10.6	Musical semantics	181
10.6.1	Neural correlates	181
10.6.2	Propositional semantics	182
10.6.3	Communication vs. expression	182
10.6.4	Meaning emerging from large-scale relations	183
10.6.5	Further theoretical accounts	184
11	Music and Action	186
11.1	Perception–action mediation	186
11.2	ERP correlates of music production	189

12	Emotion	203
12.1	What are ‘musical emotions’?	204
12.2	Emotional responses to music – underlying mechanisms	207
12.3	From social contact to spirituality – The Seven Cs	208
12.4	Emotional responses to music – underlying principles	212
12.5	Musical expectancies and emotional responses	216
12.5.1	The tension-arch	218
12.6	Limbic and paralimbic correlates of music-evoked emotions	219
12.6.1	Major–minor and happy–sad music	225
12.6.2	Music-evoked dopaminergic neural activity	226
12.6.3	Music and the hippocampus	227
12.6.4	Parahippocampal gyrus	231
12.6.5	A network comprising hippocampus, parahippocampal gyrus, and temporal poles	232
12.6.6	Effects of music on insular and anterior cingulate cortex activity	232
12.7	Electrophysiological effects of music-evoked emotions	233
12.8	Time course of emotion	234
12.9	Salutary effects of music making	235
13	Concluding Remarks and Summary	241
13.1	Music and language	241
13.2	The music-language continuum	244
13.3	Summary of the theory	249
13.4	Summary of open questions	258
	References	267
	Index	303

# Preface

Music is part of human nature. Every human culture that we know about has music, suggesting that, throughout human history, people have played and enjoyed music. The oldest musical instruments discovered so far are around 30 000 to 40 000 years old (flutes made of vulture bones, found in the cave *Hohle Fels* in Geissenklösterle near Ulm in Southern Germany).<sup>1</sup> However, it is highly likely that already the first individuals belonging to the species *homo sapiens* (about 100 000 to 200 000 years ago) made musical instruments such as drums and flutes, and that they made music cooperatively together in groups. It is believed by some that music-making promoted and supported social functions such as communication, cooperation and social cohesion,<sup>2</sup> and that the human musical abilities played a key phylogenetic role in the evolution of language.<sup>3</sup> However, the adaptive function of music for human evolution remains controversial (and speculative). Nevertheless, with regard to human ontogenesis, we now know that newborns (who do not yet understand the syntax and semantics of words and sentences) are equipped with musical abilities, such as the ability to detect changes of musical rhythms, pitch intervals, and tonal keys.<sup>4</sup> By virtue of these abilities, newborn infants are able to decode acoustic features of voices and prosodic features of languages.<sup>5</sup> Thus, infants' first steps into language are based on prosodic information (that is, on the musical aspects of speech). Moreover, musical communication in

---

<sup>1</sup> Conard *et al.* (2009)

<sup>2</sup> Cross & Morley (2008), Koelsch *et al.* (2010a)

<sup>3</sup> Wallin *et al.* (2000)

<sup>4</sup> Winkler *et al.* (2009b) Stefanics *et al.* (2007), Perani *et al.* (2010)

<sup>5</sup> Moon *et al.* (1993)

early childhood (such as parental singing) appears to play an important role in the emotional, cognitive, and social development of children.<sup>6</sup>

Listening to music, and music making, engages a large array of psychological processes, including perception and multimodal integration, attention, learning and memory, syntactic processing and processing of meaning information, action, emotion, and social cognition. This richness makes music an ideal tool to investigate human psychology and the workings of the human brain: Music psychology inherently covers, and connects, the different disciplines of psychology (such as perception, attention, memory, language, action, emotion, etc.), and is special in that it can combine these different disciplines in coherent, integrative frameworks of both theory and research. This makes music psychology *the* fundamental discipline of psychology.

The neuroscience of music is music psychology's tool to understanding the human brain. During the last few years, neuroscientists have increasingly used this tool, which has led to significant contributions to social, cognitive, and affective neuroscience. The aim of this book is to inform readers about the current state of knowledge in several fields of the neuroscience of music, and to synthesize this knowledge, along with the concepts and principles developed in this book, into a new theory of music psychology.

The first part of this book consists of seven introductory chapters. Their main contents are identical to those of a 'first edition' of this book (the publication of my PhD thesis), but I have updated the chapters with regard to scientific developments in the different areas. These chapters introduce the ear and hearing, a few music-theoretical concepts, perception of pitch and harmony, neurophysiological mechanisms underlying the generation of electric brain potentials, components of the event-related brain potential (ERP), the history of electrophysiological studies investigating music processing, and functional neuroimaging techniques. The purpose of these introductory chapters is to provide individuals from different disciplines with essential knowledge about the neuroscientific, music-theoretical, and music-psychological concepts required to understand the second part of the book (so that individuals without background knowledge in either of these areas can nevertheless understand the second part). I confined the scope of these chapters to those contents that are relevant to understanding the second part, rather than providing exhaustive accounts of each area. Scholars already familiar with those areas can easily begin right away with the second part.

The second part begins with a chapter on a model of music perception (Chapter 8). This model serves as a theoretical basis for processes and concepts developed in the subsequent chapters, and thus as a basis for the construction of the theory of music psychology introduced in this book. The chapter is followed by a chapter on music-syntactic processing (Chapter 9). In that chapter, I first tease apart different cognitive operations underlying music-syntactic processing.

---

<sup>6</sup> Trehub (2003)

In particular, I advocate differentiating between: (a) processes that do not require (long-term) knowledge, (b) processes that are based on long-term knowledge and involve processing of local, but not long-distance, dependencies, and (c) processing of hierarchically organized structures (including long-distance dependencies). Then, I provide a detailed account on studies investigating music-syntactic processing using the *early right anterior negativity* (ERAN). One conclusion of these studies is the *Syntactic Equivalence Hypothesis* which states that there exist cognitive operations (and neural populations mediating these operations) that are required for music-syntactic, language-syntactic, action-syntactic, as well as mathematical-syntactic processing, and that are neither involved in the processing of acoustic deviance, nor in the processing of semantic information.

Chapter 10 deals with music-semantic processing. Here I attempt to tease apart the different ways in which music can either communicate meaning, or evoke processes that have meaning for the listener. In particular, I differentiate between *extra-musical* meaning (emerging from iconic, indexical, and symbolic sign quality), *intra-musical* meaning (emerging from structural relations between musical elements), and *musicogenic* meaning (emerging from music-related physical activity, emotional responses, and personality-related responses). One conclusion is that processing of extra-musical meaning is reflected electrically in the N400 component of the ERP, and processing of intra-musical meaning in the N5 component. With regard to musicogenic meaning, a further conclusion is that music can evoke sensations which, *before* they are ‘reconfigured’ into words, bear greater inter-individual correspondence than the words that an individual uses to describe these sensations. In this sense, music has the advantage of defining a sensation without this definition being biased by the use of words. I refer to this musicogenic meaning quality as *a priori musical meaning*.

Chapter 11 deals with neural correlates of music and action. The first part of that chapter reviews studies investigating premotor processes evoked by listening to music. The second part reviews studies investigating action with ERPs. These studies investigated piano playing in expert pianists, with a particular focus on (a) ERP correlates of errors that the pianists made during playing, and (b) processing of false feedback (while playing a correct note). Particularly with regard to its second part, this chapter is relatively short, due to the fact that only few neuroscientific studies are yet available in this area. However, I regard the topic of music and action as so important for the neuroscience of music, that I felt that something was missing without this chapter.

Chapter 12 is a chapter on music-evoked emotions and their neural correlates. It first provides theoretical considerations about principles underlying the evocation of emotion with music. These principles are not confined to music, but can be extrapolated to emotion psychology in general. I also elaborate on several social functions that are engaged when making music in a group. One proposition is that music is special in that it can activate all of these social functions at the same time. Engaging in these functions fulfils human needs, and can, therefore, evoke strong emotions. Then, a detailed overview of functional neuroimaging studies

investigating emotion with music is provided. These studies show that music-evoked emotions can modulate activity in virtually all so-called limbic/paralimbic brain structures. This indicates, in my view, that music-evoked emotions touch the core of evolutionarily adaptive neuroaffective mechanisms, and reflects that music satisfies basic human needs. I also argue that experiences of fun and reward have different neural correlates than experiences of joy, happiness, and love. With regard to the latter emotions, I endorse the hypothesis that they are generated in the hippocampus (and that, on a more general level, the hippocampus generates tender positive emotions related to social attachments). In the final section of that chapter, I present a framework on salutary effects of music making. Due to the scarcity of studies, that framework is thought of as a basis for further research in this area.

In the final chapter, I first provide a concluding account on ‘music’ and ‘language’. I argue that there is no design feature that distinctly separates music and language, and that even those design features that are more prominent in either language or music also have a transitional zone into the respective other domain. Therefore, the use of the words ‘music’ and ‘language’ seems adequate for our everyday language, but for scientific use I suggest the term *music-language-continuum*.

Then, the different processes and concepts developed in the preceding chapters are summarized, and synthesized into a theory of music perception. Thus, readers with very limited time can skip to page 201 and read only Section 13.3, for these few pages contain the essence of the book. In the final section, the research questions raised in the previous chapters are summarized. That summary is meant as a catalogue of research questions that I find most important with regard to the topics dealt with in the second part of this book. This catalogue is also meant to provide interested students and scientists who are new to the field with possible starting points for research.

The theory developed in this book is based on the model of music perception described in Chapter 8; that model describes seven stages, or dimensions, of music perception. The principles underlying these dimensions are regarded here as so fundamental for music psychology (and psychology in general), that processes and concepts of other domains (such as music perception, syntactic processing, musical meaning, action, emotion, etc.) were developed and conceptualized in such a way that they correspond to the dimensions of music perception.

This led to a theory that integrates different domains (such as music, language, action, emotion, etc.) in a common framework, implying numerous shared processes and similarities, rather than treating ‘language’, ‘music’, ‘action’, and ‘emotion’ as isolated domains.<sup>7</sup> That is, different to what is nowadays common in psychology and neuroscience, namely doing research in a particular domain without much regard to other domains, the music-psychological approach taken in

---

<sup>7</sup> See also Siebel *et al.* (1990).

this book aims at bringing different domains together, and integrating them both theoretically and empirically into a coherent theory. In this regard, notably, this book is about understanding human psychology and the human brain (it is *not* about understanding music, although knowledge about how music is processed in the brain can open new perspectives for the experience of music). In my view, we do not need neuroscience to explain, or understand music (every child can understand music, and Bach obviously managed to write his music without any brain scanner). However, I do believe that we need music to understand the brain, and that our understanding of the human brain will remain incomplete unless we have a thorough knowledge about how the brain processes music.

Many of my friends and colleagues contributed to this book through valuable discussions, helpful comments, and numerous corrections (in alphabetical order): Matthias Bertsch, Rebecca Chambers-Lepping, Ian Cross, Philipp Engel, Thomas Fritz, Thomas Gunter, Thomas Hillecke, Sebastian Jentschke, Carol Lynne Krumhansl, Moritz Lehne, Eun-Jeong Lee, Giacomo Novembre, Burkhard Maess, Clemens Maidhof, Karsten Müller, Jaak Panksepp, Uli Reich, Tony Robertson, Martin Rohrmeier, María Herrojo Ruiz, Daniela Sammler, Klaus Scherer, Walter Alfred Siebel, Stavros Skouras, and Kurt Steinmetzger. Aleksandra Gulka contributed by obtaining the reprint permissions of figures. It is with great joy that I see this book now finalized and in its entirety, and I hope that many readers will enjoy reading this book.

Stefan Koelsch  
Leipzig, Germany





# Part I

## Introductory Chapters



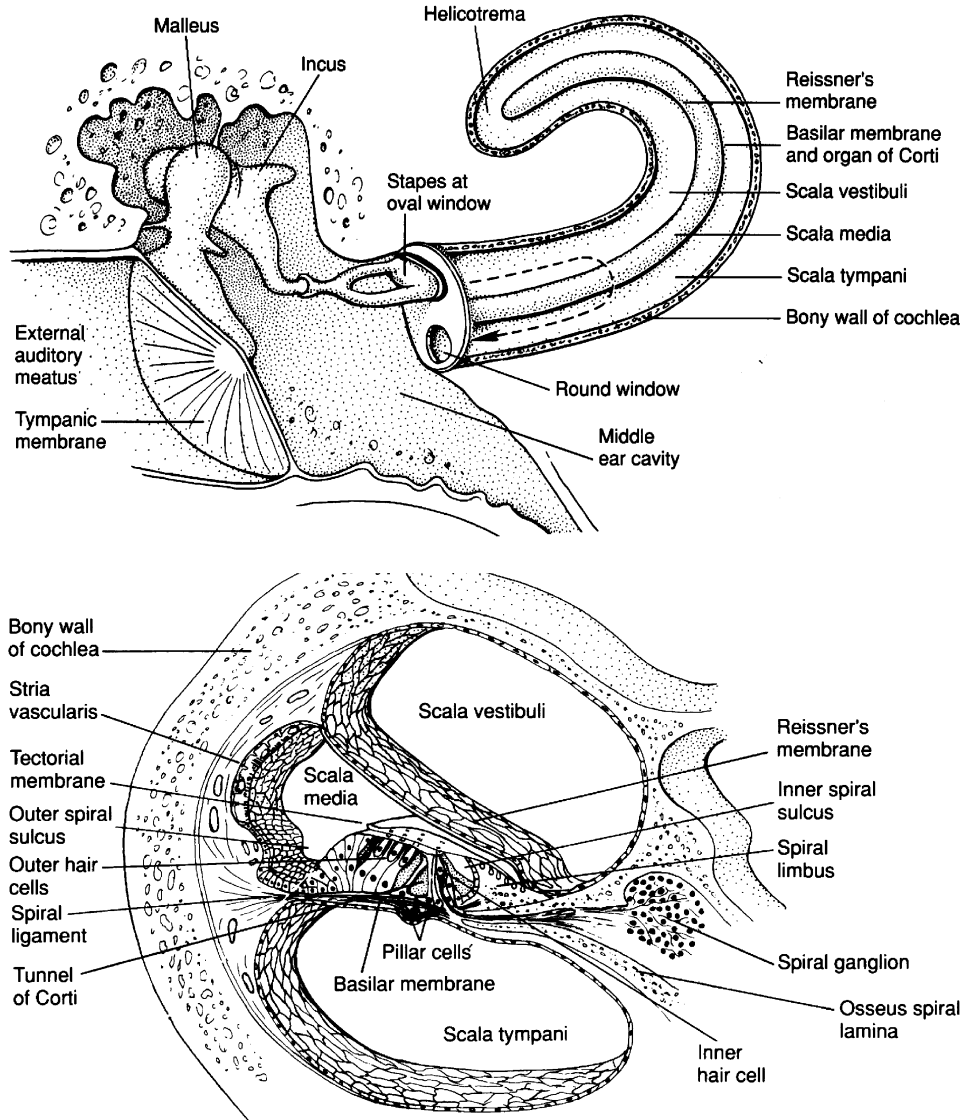
# 1

## 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



**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)<sup>1</sup> 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).

---

<sup>1</sup> 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.

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).

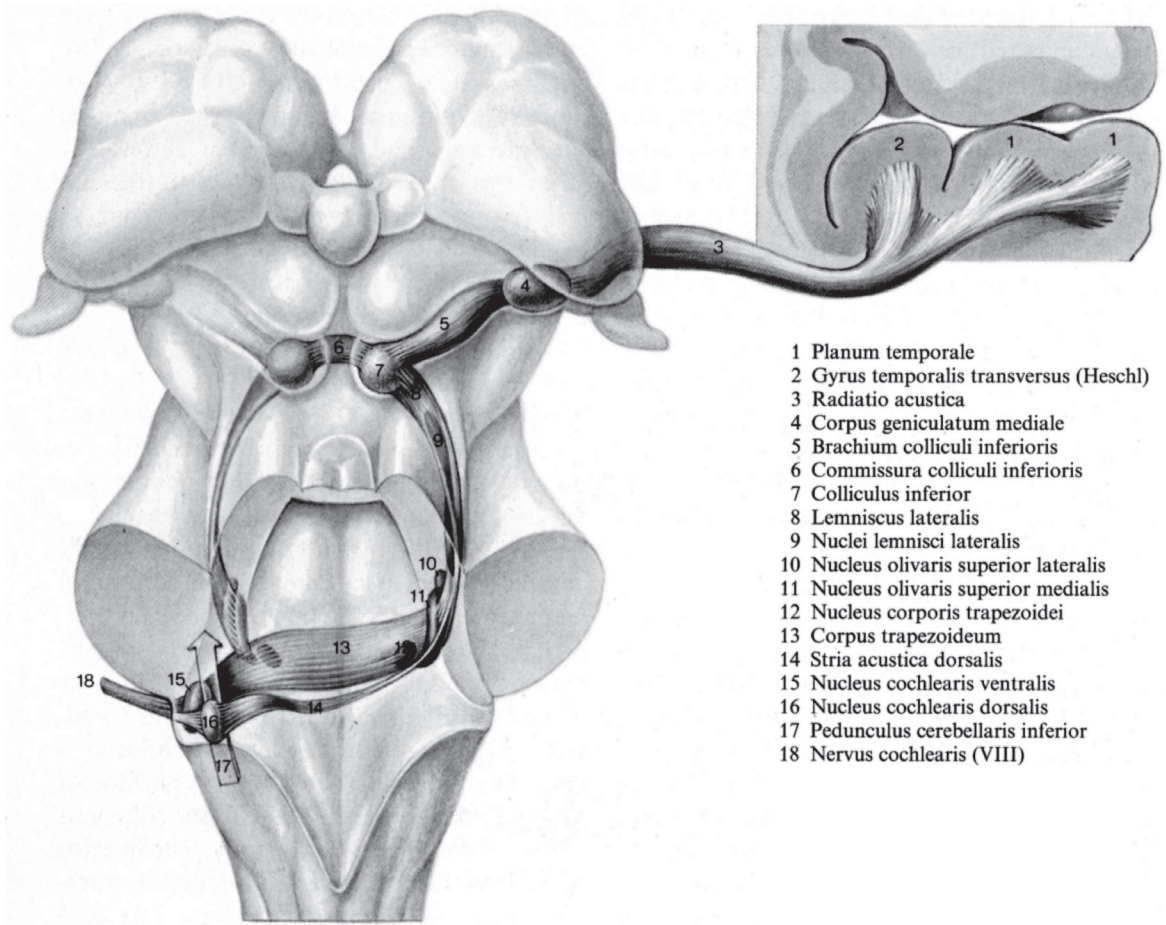
## 1.2 Auditory brainstem and thalamus

The cochlear nerve enters the central nervous system in the brain stem (cranial nerve VIII).<sup>2</sup> 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

---

<sup>2</sup> 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).<sup>3</sup>

### 1.3 Place and time information

The tonotopic excitation of the basilar membrane (BM),<sup>4</sup> 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*<sup>5</sup> (for details about the *place theory* and *temporal theory* see, e.g., Moore, 2008).

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

<sup>4</sup> 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.

<sup>5</sup> 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*,<sup>6</sup> 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);<sup>7</sup> 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),<sup>8</sup> 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*<sup>9</sup> (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

<sup>6</sup> The German term for ‘residue pitch’ is ‘Residualton’.

<sup>7</sup> 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.

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

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

down four times in one second). When the beats have higher frequencies (above  $\sim 20$  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*.<sup>10</sup> 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.<sup>11</sup> 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).<sup>12</sup> 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.

<sup>10</sup> (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.

<sup>11</sup> 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).

<sup>12</sup> (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.<sup>13</sup>

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.<sup>14</sup> 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’).<sup>15</sup> Both are characterized by the two physical correlates of timbre: *Spectrum envelope* (i.e.,

---

<sup>13</sup> 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).

<sup>14</sup> 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.

<sup>15</sup> 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).<sup>16</sup> 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).<sup>17,18</sup>

Figure 1.3 shows these regions and their connectivity according to the nomenclature introduced by Kaas & Hackett (2000).<sup>19</sup> 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

---

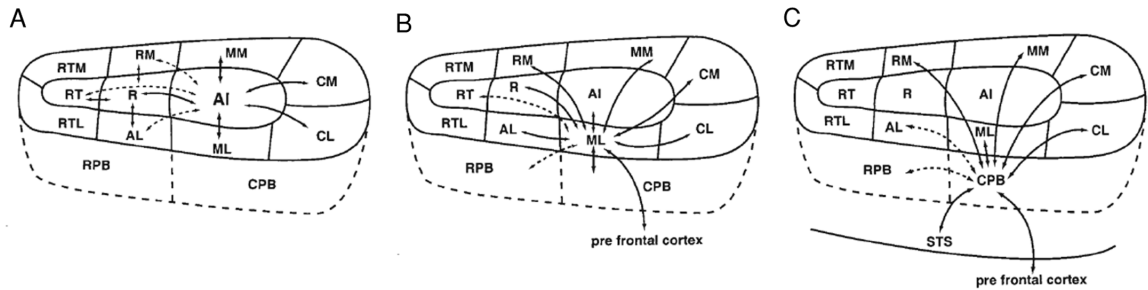
quality. There are also many examples of timbre differences in speech. For example, two vowels spoken with the same loudness and same pitch differ from one another in timbre.

<sup>16</sup> 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.'

<sup>17</sup> 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).

<sup>18</sup> 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.

<sup>19</sup> 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 rostrotemporal 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 (RPM) 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 rostrotemporal area; RTM, medial rostrotemporal 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.<sup>20</sup> 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.<sup>21</sup>

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.<sup>22</sup> That is, different mechanisms contribute to the perception of residue pitch on at least

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

<sup>21</sup> 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).

<sup>22</sup> 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.