OXFORD

AUDITORY TEMPORAL PROCESSING
AND ITS DISORDERS

Jos J. Eggermont

Auditory Temporal Processing and its Disorders

Auditory Temporal Processing and its Disorders

Jos J. Eggermont

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First Edition published in 2015

Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2014955853

ISBN 978–0–19–871909–0

Printed and bound by
CPI Group (UK) Ltd, Croydon, CR0 4YY

Preface

Sound reflects the dynamic spectrum of pressure fluctuations in the air. The auditory system correlates of spectral analysis and representation are well known and studied with respect to adjustments of frequency-specific hearing loss with hearing aids. The temporal aspects of sound as analyzed and represented in the auditory system have also been thoroughly studied, whereas the disorders of temporal processing have only recently become recognized. They cannot be amended with hearing aids, as their problem is typical of the central nervous system, albeit that the origin might also lie frequently in the cochlea and auditory nerve. Temporal mechanisms underlie pitch perception, which allows the enjoyment of music, and are required to make a distinction between voiced and unvoiced consonants. Temporal processing disorders result in the inability or great difficulty to understand speech, to communicate, and to fit into society.

Whereas frequency-specific hearing problems are exclusively an auditory perceptual problem, temporal processing problems are frequently not limited to the auditory system but may reflect a general neurological disorder. The exquisite temporal resolution of the auditory system may well allow for early detection in this domain and also is reflected in production and understanding of speech, the sound of choice for human communication. Speech is spoken language and problems with language acquisition either phonologically or orthographically reflect themselves in a trio of disorders, such as SLI, dyslexia, and auditory processing disorder, that may largely reflect similar underlying pathology.

Understanding, diagnosing, and rehabilitation of the various auditory temporal processing disorders requires a good understanding of the basic temporal processing that is carried out by the auditory system, peripheral as well as central. We will see that forward masking, gap detection, and temporal modulation detection are all based on and can largely be understood on the basis of neural firing-rate adaptation. These are the bottom-up mechanisms, but they are not sufficient for understanding the clinical problems. This requires incorporating attention and general cognitive mechanisms of communication, and integrating these top-down processes with the bottom-up ones.

This book gives first of all an account of the basic temporal aspects of sound: onsets, echoes, adaptation to continuing sound, the effects of one sound upon another in the form of forward masking, and detection of a gap between them. Amplitude modulations of a sound, coincidences in these aspects for different frequencies that make up a complex sound and their role in sound localization allow for the ability to keep up a conversation at a cocktail party. Rhythm and pitch have formed the basis for theories of hearing for well over two centuries, and still provide the basis for meaningful employment for hundreds of auditory researchers. This knowledge is also important to the design of ever better sound processors for cochlear implants.

If we take the temporal processing ability of normal hearing young adults as our standard, then infants, children, and adolescents as well as the elderly have an auditory temporal processing deficit. We will describe this in some detail, as it potentially confounds the diagnosis of auditory processing disorders. Real auditory temporal processing problems occur when the ability of auditory nerve (neuropathy) and central nerve tracts (multiple sclerosis) to conduct rapidly the activity produced by sound is compromised by deficits in the insulation of the individual nerve fibers. Putative temporal processing problems such as in gap detection and amplitude modulation detection, but also in being able to hear small differences in the timing of sounds, have been claimed to underlie dyslexia and SLI as well as auditory processing disorder. Others claim that all these problems relate to cognitive processing problems. General neurological and psychiatric disorders present with distortions in time perception and timed performance (e.g., schizophrenia, ASD, and epilepsy), but not only in the perception of sound. We will investigate the role of neural synchrony deficits underlying auditory temporal processing problems in these disorders. Finally, I will describe the role of multimodal integration in several auditory temporal processing disorders, and present a comparative overview of the various temporal processing deficits in all the discussed disorders. I had to be selective in my description of the many contributions to this vast field of auditory, audiological, and neurological research. If you are not listed among the more than one thousand references, I apologize for my biases and potential oversight.

Most of this work was supported by Alberta Innovates-Health Solutions, the Natural Sciences and Engineering Research Council of Canada, and the Campbell McLaurin Chair for Hearing Deficiencies.

I wholeheartedly thank Dennis Phillips and Martin Pienkowski who took considerable time and effort to read through the entire manuscript and provide numerous insightful comments. My wife Mary—translator, writer, and editor—read various drafts and corrected grammatical and stylistic blemishes. I also acknowledge my acquisition editor at OUP, Charlotte Green, for guiding me through the writing process. They all made this a better read.

Calgary, August 2014

Contents

List of Abbreviations

A1	primary auditory cortex
AAF	anterior auditory field
ABR	auditory brainstem response
ACF	autocorrelation function
AD	axial diffusivity; after discharge
AEF	auditory magnetic fields
AEP	auditory-evoked potential
AF	audio-first
AFM	auditory field map
AHP	after-hyperpolarization
AI	cat primary auditory cortex
AII	cat secondary auditory cortex
AL	anterolateral
ALSR	average localized synchronized rate
AM	amplitude modulation
AN	auditory nerve
ANF	auditory nerve fiber
ANP	auditory neuropathy
AP	auditory processing
APD	auditory processing disorder
ART	auditory repetition test
AS	Aspberger syndrome
ASD	autism spectrum disorder
ASSR	auditory steady-state response
AV	audio-visual
AVCN	antero-ventral cochlear nucleus
BD	best delay
BF	best frequency
BIC	brachium of the inferior colliculus
BM	backward masking
BMF	best modulation frequency
BRB-N	brief repeatable battery-neuropsychology
CANS	central auditory nervous system
CAP	compound action potential
CAPD	central auditory processing disorder
CB	critical band
CC	corpus callosum
CDT	cubic difference tone
CF	characteristic frequency
CHL	conductive hearing loss
CL	caudolateral belt area
CM	cochlear microphonic; caudomedial belt area
CMF	corner modulation frequency
CN	cochlear nucleus
CNV	contingent negative variation
CSD	current source density
CV	coefficient of variation; consonant–vowel
DCN	dorsal cochlear nucleus
DL	difference limen
DNLL	dorsal nucleus of the lateral lemniscus
DPOAE	distortion-product otoacoustic emission
DPST	duration pattern sequence test
DT	difference tone
DTI	diffusion tensor imaging
ECochG	electrocochleogram
EDSS	Expanded Disability Status Scale
EEG	electroencephalogram
EPSC	excitatory postsynaptic current
EPSP	excitatory postsynaptic potential
ERP	event-related brain potential
F_0	fundamental frequency
FA	fractional anisotropy
FFR	frequency following response
FM	frequency modulation
fMRI	functional magnetic resonance imaging
FSL	first-spike latency
FTC	frequency-tuning curve
GBC	globular bushy cell
GIN	gaps-in-noise
GWN	Gaussian White Noise
HF	high frequency
HFA	high-functioning autists

HG	Heschl's gyrus
HI	hearing impaired
HINT	hearing in noise test
Hz	Hertz
IC	inferior colliculus
ICC	central nucleus of the IC
ICX	external (cortical) nucleus of the IC
Ig	granular insula
IHC	inner hair cell
ILD	inter-aural level difference
IPD	inter-aural phase difference
IRN	iterated rippled noise
ISI	inter-spike interval; inter-stimulus interval
ITD	inter-aural time difference
JND	just-noticeable difference
LFP	local field potential
LI	learning impaired; learning impairments
LIF	leaky integrate-and-fire
LL	lateral lemniscus
LLR	long-latency response
LNTB	lateral nucleus of the trapezoid body
LP	learning problem
LSO	lateral superior olive
LTD	long-term depression
LTP	long-term potentiation
MD	mean diffusivity
MEG	magnetoencephalography
MET	mechano-electrical transduction
MF	modulation frequency
MGB	medial geniculate body
MGBv	ventral medial geniculate body
MGC	see MGB
MGT	minimum gap thresholds
ML	middle lateral
MLD	masking level difference
MLR	middle latency response
MMN	mismatch negativity
MMNp	putative MMN
MNTB	medial nucleus of the trapezoid body
MRI	magnetic resonance imaging
MS	multiple sclerosis
MS-CI	cognitively impaired MS patients
MS-CP	cognitively preserved MS patients
MSN	multisensory nuclei in posterior thalamus
MSO	medial superior olive
MTF	modulation transfer function
MU	multiunit
MUA	multiunit activity
NC	normal control
NH	normal hearing
nHL	normal hearing level
NL	nucleus laminaris
NSRAN	non-syndromic recessive auditory neuropathy
NSRHL	non-syndromic recessive hearing loss
ORN	object-related negativity
OTOF	otoferlin
PAC	primary auditory cortex (humans)
PAF	posterior auditory field
PLF	phase-locking factor
PLN	primary like with notch
PLS	paradoxical latency shift
PO	posterior group of thalamic nuclei
PR	presentation rate
PRR	pulse repetition rate
PSP	postsynaptic potential
PSTH	post-stimulus-time histogram
PT	planum temporale
PV	parietoventral area
PVCN	postero-ventral cochlear nucleus
R	rostral core area
RD	radial diffusivity; reading disability
Ri	retroinsular area
RM	rostromedial belt area
rMTF	rate-based modulation transfer function
RRP	ready releasable pool
RRTF	repetition rate transfer function
RT	reaction time; response time
S2	second somatosensory area
SAC	secondary auditory cortex (humans)
SAM	sinusoidal amplitude modulation
SBC	spherical bushy cell
SD	standard deviation
SEM	standard error of the mean

SFR	spontaneous firing rate
SGN	spiral ganglion neuron
SJ	synchrony judgment
SL	sensation level
SLI	specific language impairment
sMTF	synchronization modulation transfer function
SNHL	sensorineural hearing loss
SNP	single-nucleotide polymorphism
SNR	signal-to-noise ratio
SOA	stimulus-onset asynchrony
SOC	superior olivary complex
SP	summating potential
SPL	sound pressure level
SRI	specific reading impairment
SRT	speech reception threshold
SSA	stimulus-specific adaptation
STG	superior temporal gyrus
STM	short-term memory
STRF	spectro-temporal receptive field
SU	single units
susAPD	suspected auditory processing disorder
TC	thalamo-cortical
TD	typical development
TLE	temporal lobe epilepsy
tMTF	temporal modulation transfer function
TOJ	temporal order judgment
TPO	temporal parietal occipital area
Tpt	temporal parietotemporal area
TW	temporal-window
VAS	virtual acoustic space
VBM	voxel-based morphometry
VCN	ventral cochlear nucleus
VF	video-first
VNLL	ventral nucleus of the lateral lemniscus
VNTB	ventral nucleus of the trapezoid body
VOT	voice-onset-time
VP	ventroposterior complex (of the thalamus)
VS	vector strength
VSRF	virtual space receptive field
WM	white matter

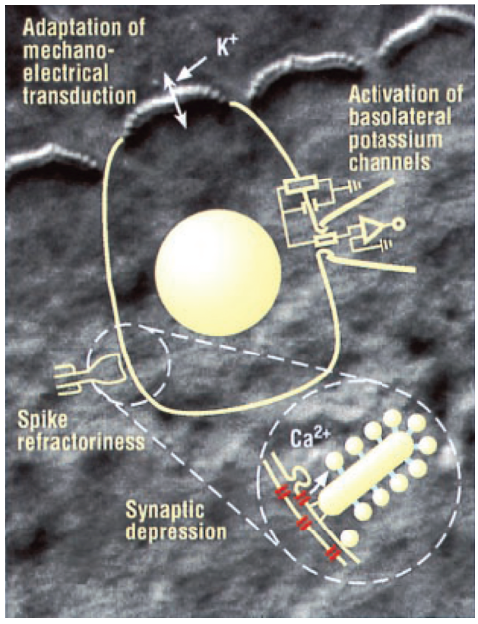

Plate 1 Exocytosis recorded by membrane capacitance measurements from single IHCs. Simplified representation of an IHC, the measuring configuration, and an afferent IHC synapse (**Inset**) overlaid onto a Nomarski image of the explanted mouse organ of Corti (stereocilia of the row of IHCs are shown at the top of the image; cell bodies are covered by supporting cells). Previously suggested mechanisms of adaptation are indicated according to their localizations. Synaptic depression could be due to Ca^{2+} current inactivation, vesicle depletion, or postsynaptic glutamate receptor desensitization. Reproduced from Tobias Moser and Dirk Beutner, Kinetics of Exocytosis and Endocytosis at the Cochlear Inner Hair Cell Afferent Synapse of the Mouse, *PNAS*, 97(2), pp. 883–8, Figure 1c, doi: 10.1073/pnas.97.2.883, © The National Academy of Sciences, 2000 (See Figure 3.6).

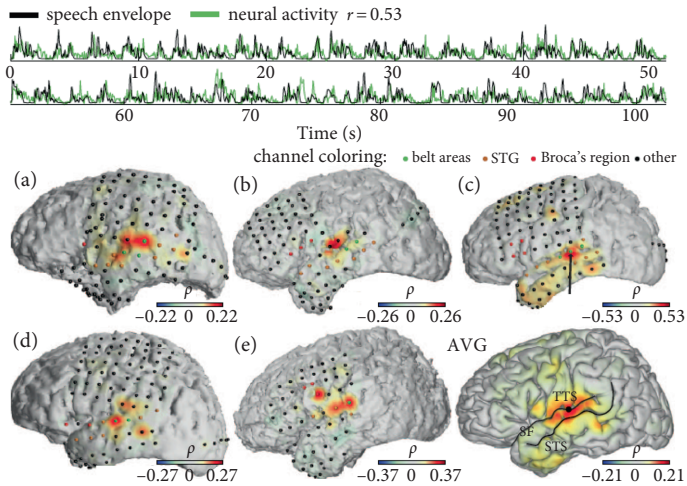

Plate 2 Top. High gamma activity in human auditory cortex tracks the envelope of speech. Black: Time course of the speech envelope. Green: High gamma activity recorded by a channel positioned in the belt areas in subject C (see Figure 3.5C bottom) while the subject listened to a narrated story. For the visualization purpose of this figure, the magnitude of the neural signal was scaled to the magnitude of the envelope signal. The Spearman correlation between the two signals is $\rho = 0.53$. Bottom. Neural tracking of speech envelope at each recording site in each subject. Color hue (see colorbars) gives r at each channel for the individual subjects (A–E), and for the subject average (AVG). Individual channels implanted in each subject are shown in green (belt areas), orange (STG), red (Broca's region), or black (other regions). In subject C, the arrow points to the channel for which the tracking effect is shown in Figure 5.3 top. Reproduced from Jan Kubanek, Peter Brunner, Aysegul Gunduz, David Poeppel, and Gerwin Schalk, The Tracking of Speech Envelope in the Human Cortex, *PLOS One*, figures 1 and 5, doi:10.1371/journal.pone.0053398 © The Authors, 2013. This figure is licensed under the Creative Commons Attribution 4.0 Generic License (See Figure 5.3).

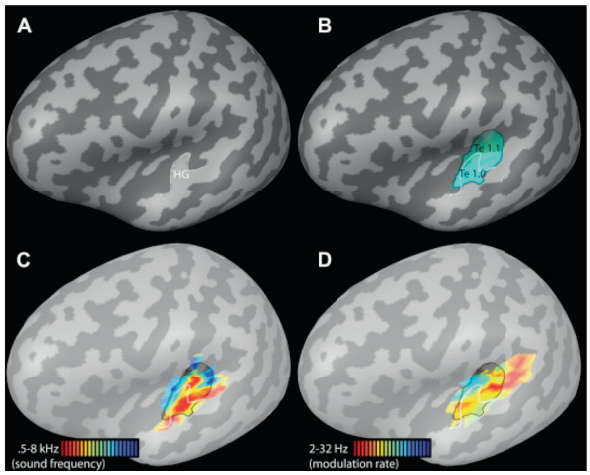

Plate 3 Group-average maps of tonotopic and modulation-rate representations for the left hemisphere. Group averages were obtained after cortex-based alignment of the individual hemispheres to achieve maximal anatomical correspondence of the observed activations across subjects and then rendered onto the reconstructed cortical surface of the Colin 27 standard brain (A, with HG outlined in white). In addition, cytoarchitectonic probability maps of auditory core regions (areas Te1.0 in bluish colors and Te1.1 in greenish colors) are displayed to allow for a comparison of anatomical (B) and functional measures ((C), (D)). (C) displays the group best-frequency (tonotopic) maps, with red colors representing low spectral frequencies, and blue for high frequencies (FDR <0.01). (D) Shows the best-modulation rate maps, with red colors representing slow temporal modulation rates, and blue for high rates of amplitude modulation (FDR <0.01). Reprinted from *Cortex*, 49 (10), Marcus Herdener, Fabrizio Esposito, Klaus Scheffler, Peter Schneider, Nikos K. Logothetis, Kamil Uludag, and Christoph Kayser, Spatial Representations of Temporal and Spectral Sound Cues in Human Auditory Cortex, pp. 2822–33, Copyright (2013), with permission from Elsevier (See Figure 5.4).

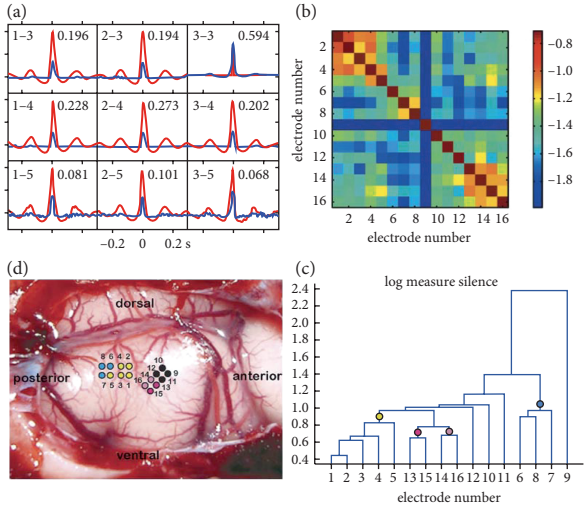

Plate 4 Clustering procedure. The basis for the clustering procedure is the pair-wise cross-correlations (A) shown here for a few pairs in their traditional form (red curves) and the corrected form (blue curves). Each panel shows the correlograms scaled on their own extremes. The time base is from—0.3 to 0.3 s. The full pair-wise correlation matrix is shown for the spikes recorded on 16 electrodes in B where the base 10 logarithm of the corrected correlation is shown (color bar indicates these values). Based on the pair-wise correlation matrix, a hierarchical clustering procedure performed on the 16 electrodes (C) results in 4 multi-electrode clusters indicated at the branching points with a colored dot. These colors are used to identify the positions of the electrodes belonging to the same cluster on the surface of the auditory cortex (D). Electrodes are numbered on the cortical surface. All electrodes are in AI. Reproduced from Jos J. Eggermont, Properties of Correlated Neural Activity Clusters in Cat Auditory Cortex Resemble those of Neural Assemblies, *Journal of Neurophysiology*, 96 (2), pp. 746–64, figure 1, doi:10.1152/jn.00059.2006 © 2006, The American Physiological Society (See Figure 7.4).

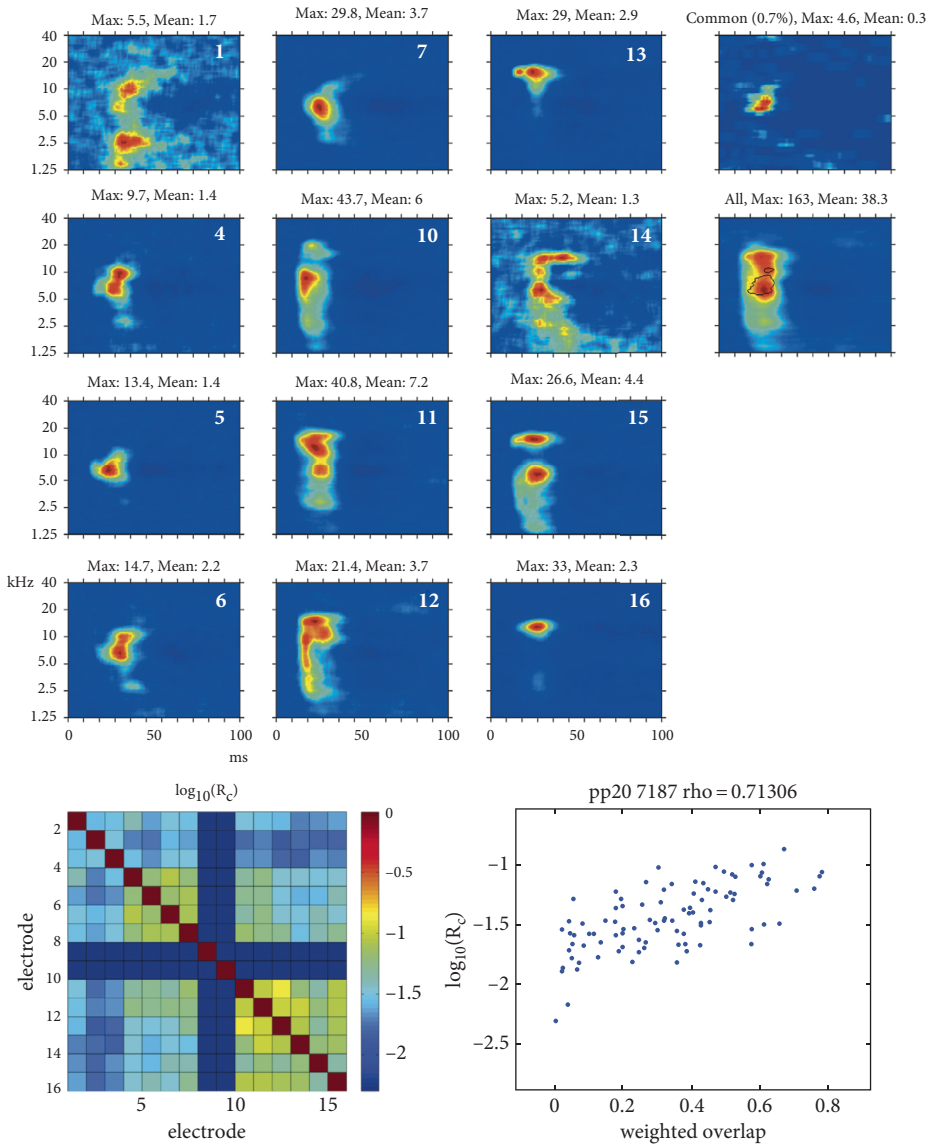

Plate 5 Example where a coincident-spike STRF, combining activity from nine electrodes in AI, is well defined and approximately equal to the overlap of the nine single-electrode STRFs. Also shown is the log of the peak cross-correlation coefficient matrix and a scattergram comparing pair-wise STRF overlap and pair-wise peak cross-correlation coefficients. The two dark areas in the correlation matrix represent electrodes that did not show correlated activity. The diagonal represents the autocorrelation coefficients, which are all equal to 1. The color bar shows the $\log_{10}(R_c)$. Reprinted from *Hearing Research*, 229 (1–2), Jos J. Eggermont, Correlated Neural Activity as the Driving Force for Functional Changes in Auditory Cortex, pp. 69–80 Copyright (2007), with permission from Elsevier (See Figure 7.6).

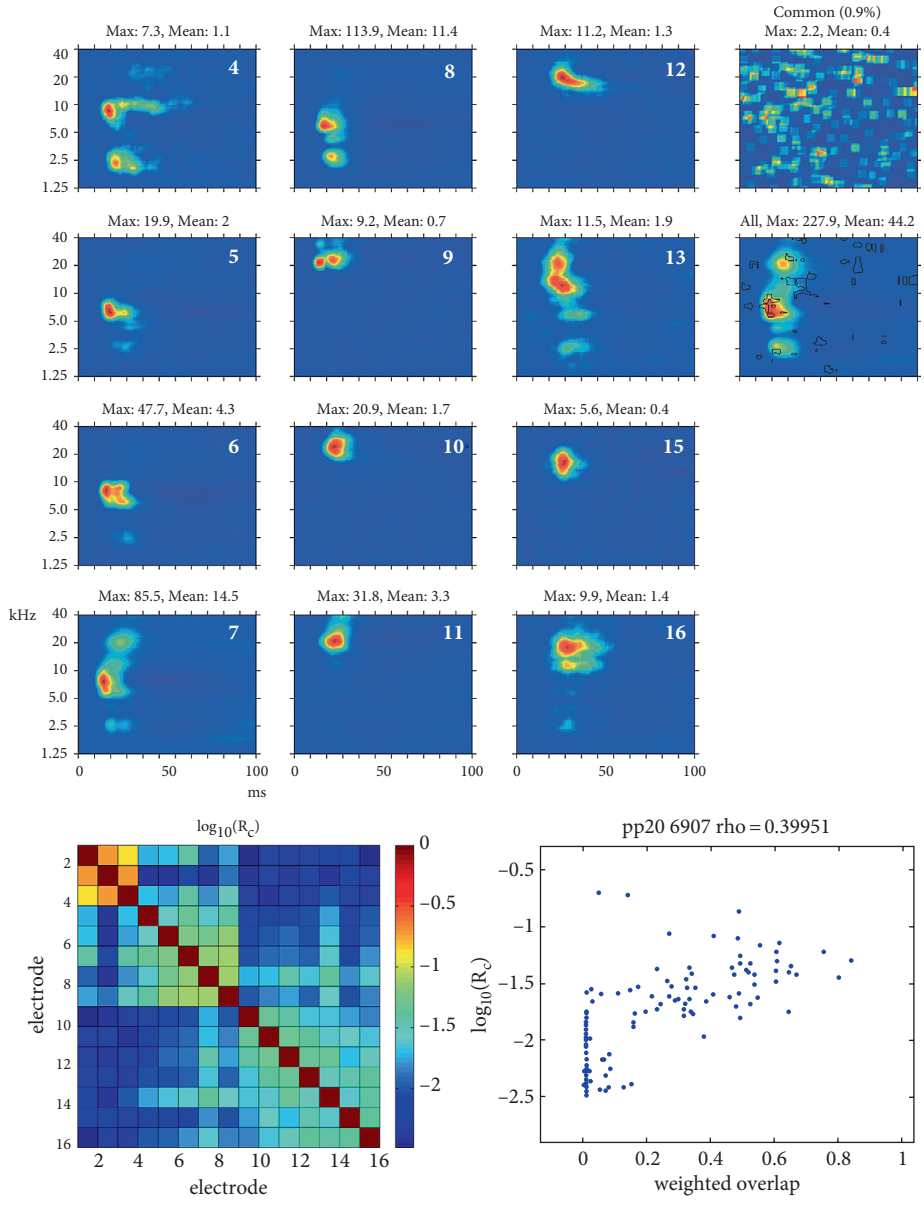

Plate 6 Example where a coincident-spike STRF (right upper panel), combining activity from nine electrodes in AI, is absent. Also shown is the log of the peak cross-correlation coefficient matrix and a scattergram comparing pair-wise STRF overlap and pair-wise peak cross-correlation coefficients. The diagonal represents the autocorrelation coefficients, which are all equal to 1. Despite the absence of a coincident-spike STRF, the pair-wise correlation matrix shows relatively high values and the nine electrodes did form a correlation cluster. The three electrodes represented in the left-hand corner of the correlation matrix formed an independent cluster with very high correlation values [$\log_{10}(R_c) > -1$] but modest STRF overlap (see scattergram). The colorbar shows the $\log_{10}(R_c)$. Reprinted from *Hearing Research*, 229 (1–2), Jos J. Eggermont, Correlated Neural Activity as the Driving Force for Functional Changes in Auditory Cortex, pp. 69–80 Copyright (2007), with permission from Elsevier (See Figure 7.7).

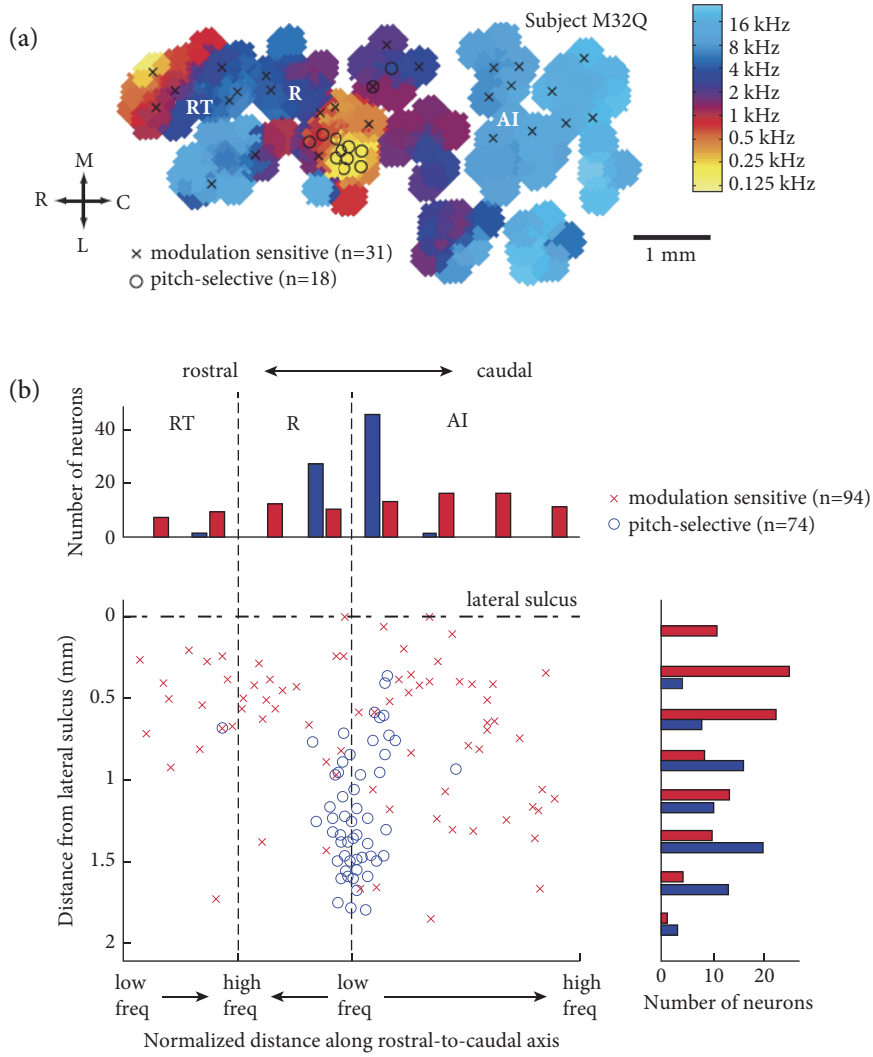

Plate 7 Spatial distribution of pitch-selective and modulation sensitive neurons. A: frequency map from 1 subject (M32Q–left hemisphere) with the location of pitch-selective neurons and modulation-sensitive neurons indicated. B: normalized cortical map of locations of pitch-selective and modulation-sensitive neurons across four subjects (four hemispheres). Reproduced from Daniel Bendor and Xiaoqin Wang, Neural Coding of Periodicity in Marmoset Auditory Cortex, *Journal of Neurophysiology*, 103 (4), pp. 1809–22, figure 9, doi:10.1152/jn.00281.2009 © 2010, The American Physiological Society (See Figure 8.6).

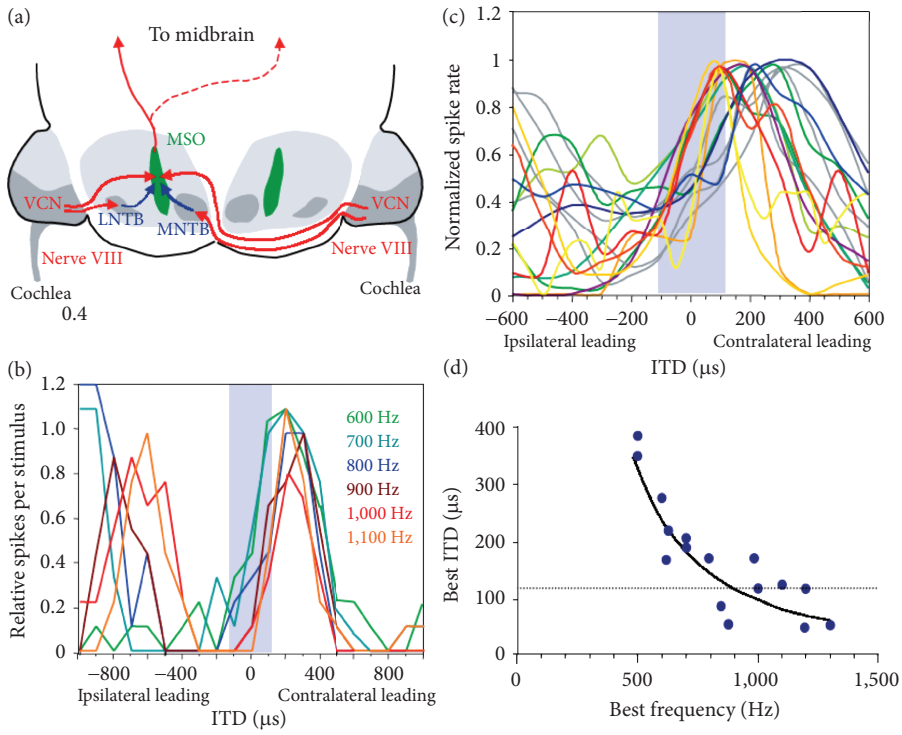

Plate 8 Left. ITD coding in the mammalian MSO. A. Innervation pattern of the MSO. Excitatory inputs (red) from VCN, and glycinergic inputs (blue) from MNTB and LNTB are shown. B. ITD function of a single gerbil MSO neuron, tested with pure tones at different frequencies. As is typical, the peak ITD is independent of stimulus frequency. The blue area indicates the physiologically relevant range for gerbils (±120 μs). C. ITD functions of gerbil MSO neurons, each tested at its best frequency, 20 dB above threshold. Note that the peaks are largely outside the physiologically relevant range (blue area). The steep slopes of the ITD functions are within the relevant area. D. Distribution of best ITDs as a function of best frequencies of the neurons. Only ITDs below the dotted line, representing the upper limit of the gerbil's physiological range of ITDs, are within the physiologically relevant range of ITDs. Reprinted by permission from Macmillan Publishers Ltd: *Nature*, 417 (6888), Antje Brand, Oliver Behrend, Torsten Marquardt, David McAlpine, and Benedikt Grothe, Precise Inhibition is Essential for Microsecond Interaural Time Difference Coding, pp. 543–7, doi:10.1038/417543a © 2002, Macmillan Publishers Ltd
(See Figure 9.2).

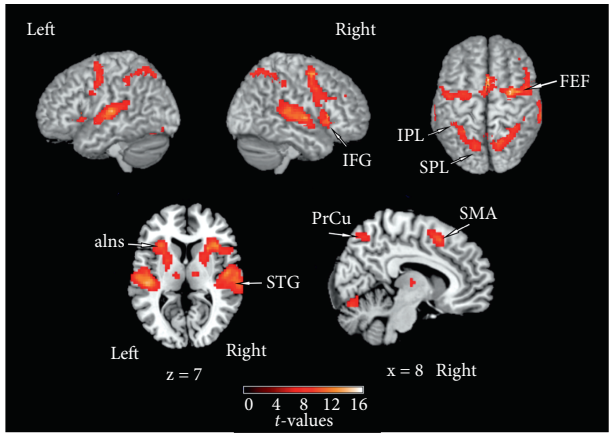

Plate 9 Activations of brain regions as revealed by the contrast of "cocktail" condition versus rest ($p < 0.05$). STG, superior temporal gyrus; FEF, frontal eye fields; IPL, inferior parietal lobule; SPL, superior parietal lobule; IFG, inferior frontal gyrus; PrCu, precuneus; aIns, anterior insula; SMA, supplementary motor area. The color code refers to t-values (see bar). Reproduced from Ida C. Zündorf, Jörg Lewald, and Hans-Otto Karnath, Neural Correlates of Sound Localization in Complex Acoustic Environments, *PLOS One*, figure 3, doi:10.1371/journal.pone.0064259 © The Authors, 2013. This figure is licensed under the Creative Commons Attribution 4.0 Generic License (See Figure 10.4).

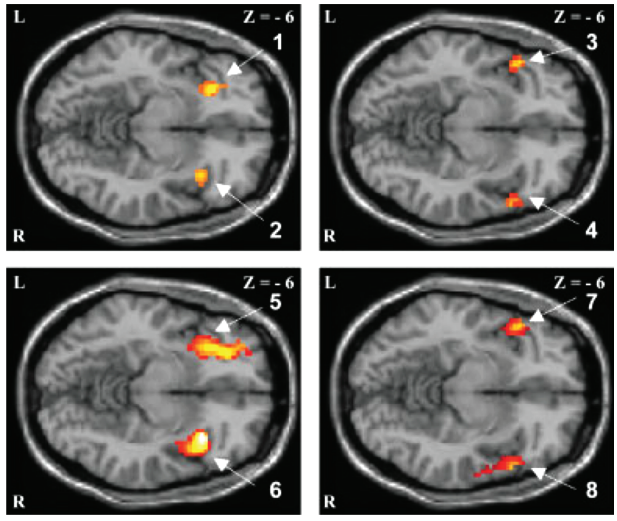

Plate 10 Subtraction analysis of the respective main effects during: (a) acoustical stimulation with clicks in healthy control subjects versus dyslexics (upper row, left illustration) and vice versa (upper row, right illustration); (b) acoustical stimulation with syllables in healthy control subjects versus dyslexics (lower row, left illustration) and vice versa (lower row, right illustration). L, left; R, right; z, distance to the intercommisural plane. Activated brain regions withe $p<0.001$: 1. Left anterior insula, 2. Right anterior insula, 3. Left rolandic operculum (frontal part), 4. Right rolandic operculum (frontal part), 5. Left anterior insula, 6. Right anterior insula, 7. Left rolandic operculum (frontal part), and 8. Right rolandic operculum (frontal part). Reproduced from Contribution of the anterior Insula to Temporal Auditory Processing Deficits in Developmental Dyslexia, Claudia Steinbrink, Hermann Ackermann, Thomas Lachmann, and Axel Riecker, *Human Brain Mapping*, 30 (8), pp. 2401–11, figure 1, © 2009 Wiley-Liss, Inc (See Figure 14.4).

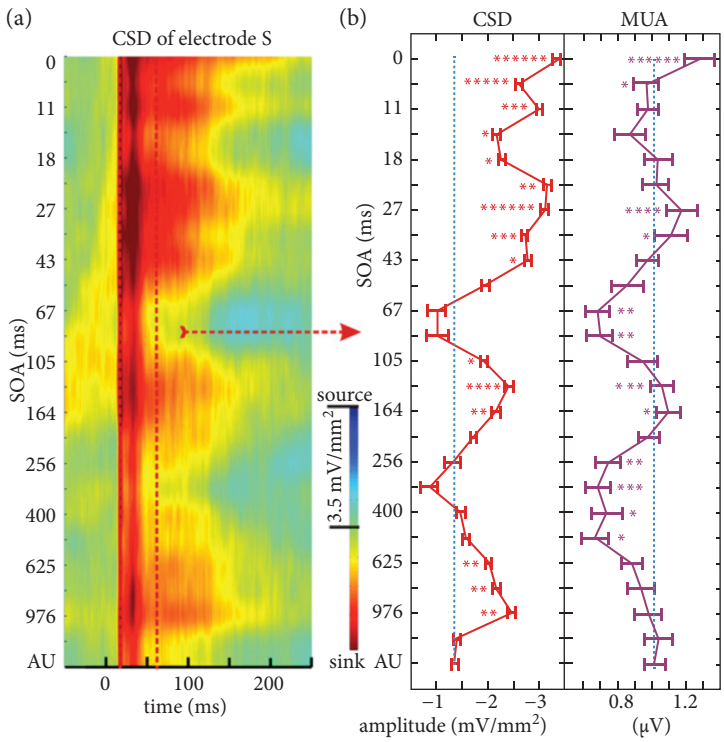

Plate 11 Effect of Somatosensory–Auditory SOA on the Supragranular Bimodal Response; (A) Color map shows the event-related CSD of the supragranular channel (S, see figure 1) in area A1 for different somatosensory–auditory SOAs. Increasing SOAs are mapped to the y-axis from top to bottom, with 0 on top corresponding to simultaneous auditory–somatosensory stimulation. AU on the bottom represents the auditory-alone condition. Red dotted lines denote the 20–60 ms time interval for which we averaged the CSD and MUA in single trials for quantitative analysis. (B) Traces show mean CSD and MUA amplitude values (*x*-axis) for the 20–60 ms auditory poststimulus time interval (error bars show standard errors) with different somatosensory–auditory SOAs (*y*-axis). Blue dotted line denotes the mean amplitude of the auditory-alone response. At a given SOA, independent-samples t tests were used for all six experiments (bimodal response amplitude in each experiment was compared with the response amplitude of the auditory-alone condition). The number of stars at a given SOA indicates how many experiments have significant differences (independent-samples t tests, $p<0.01$) in bimodal activation. Reprinted from *Neuron*, 53 (2), Peter Lakatos, Chi-Ming Chen, Monica N. O'Connell, Aimee Mills, and Charles E. Schroeder, Neuronal Oscillations and Multisensory Interaction in Primary Auditory Cortex, pp. 279–92, Copyright (2007), with permission from Elsevier (See Figure 16.1).

Chapter 1

Introduction

Speech encoding is likely the most important function of the human auditory system. The rapid changes in the spectral makeup of speech with time, e.g., as in formant transitions, are characteristic and shaped for use in communication. Speech also contains periodic elements resulting from the glottal pulses that determine the pitch of the human voice. Onsets and gaps determine the phonemes in speech. At least under optimal listening conditions, frequency information in the formants appears far less important than the temporal content of speech for recognition of phonemes and words in simple sentences (Shannon et al. 1995):

> High speech recognition performance can be achieved with only three time-varying bands of noise representing the complex spectral patterns of speech. . . . Harmonic structure of voiced speech was not present in the noise-band simulations. Despite this reduced spectral content, the temporal cues were sufficient to produce 90% correct identification of words.

1.1 Multiple aspects of auditory temporal processing

When we talk about the temporal response properties of the auditory system we typically refer to the way the timing of neural activity in the nervous system follows the various temporal aspects of sound (Rosen 1992). Temporal aspects of sound are present in the sound carrier, i.e., the carrier frequencies in the sound waveform, as well as in the sound envelope. The latter reflects the amplitude and frequency modulations of the carrier waveform, including starts, stops, and gaps. Phillips (1993b) phrased it as follows:

> By definition, sounds are physical events that are distributed in time. It follows from this that the faithful neural encoding of a sound requires that the nervous system in some way preserve or "represent" the relevant time structure of the signal in the cadence of spike discharges evoked by the sound.

Temporal processing of sound can be divided into two broad aspects: temporal integration and temporal acuity (Eddins and Green 1995). Temporal integration is the process that among others describes time-intensity trade, meaning that sounds of longer duration can be detected earlier and at lower levels. This aspect of temporal processing can be described by a low-pass filtering action with a time constant of about 0.2 s. We will see that abnormalities in this integration process play a role in some temporal processing disorders. Temporal acuity or resolution is a much faster process and plays a role in forward and backward masking, gap detection, and amplitude modulation detection, as well as in temporal order judgment (first described by Hirsh 1959; see Chapter 16)

which is related to gap detection. This process can be modeled by a high-pass filtering action, with a time constant as short as 2 ms but ranging up to 30 ms (Eddins and Green 1995). This aspect of temporal processing of sound is the topic of Chapters 3–8. It is interesting that these seemingly opposite operations may be carried out by the same neural population. Neocortical pyramidal cells may act as coincidence detectors or temporal integrators depending on the degree of synchrony among their synaptic inputs—thalamocortical as well as intracortical ones (Aertsen et al. 1994; Grande et al. 2004). High synchrony of inputs to a given pyramidal cell leads to the more efficient coincidence detection, whereas low input synchrony leads to temporal integration. Temporal integration as well as coincidence detection may lead to similar extracted features such as harmonic tuning in the auditory cortex (Chapter 7). The neurons over which temporal integration or coincidence detection is carried out may belong to clusters that exhibit stronger correlation among its member neurons compared to that with other neurons (Eggermont 2006).

1.1.1 Faithful representation of the temporal structure of sound

The temporal aspects of sound thus encompass carrier fine structure, onsets and offsets, gaps, and periodicities. The capacity of neurons in the auditory system to follow the temporal aspects of sound is limited first of all by the intrinsic properties of auditory nerve fibers (ANFs) that transmit these aspects to the central auditory nervous system. One important limitation is the refractoriness following an action potential. This prohibits any neuron to faithfully follow frequencies of more than ~1 kHz, i.e., periods of <1 ms. This frequency-following limit of individual ANFs can be partially overcome by a population code that adheres to the "volley" principle. According to this principle, individual neurons reliably phase-lock to the carrier wave but not to all periods thereof, but over time or as a population may cover all of these individual cycles (Wever 1949). Thus for a pure tone, the ensemble activity in the auditory nerve over a short period of time mimics the activity of single ANFs over a relatively long period of time. In cats, this time-average phase-locking of a single ANF or as an ensemble average in a population is significant up to 5 kHz (Johnson 1980).

1.1.1.1 Gap detection and voice-onset-time

In speech, a voiced stop consonant such as /b/ can be distinguished from its voiceless counterpart /p/ by temporal cues. For phonemes such as /ba/ and /pa/, the distinction is generally based on the length of the interval between consonant release and the onset of voicing (Figure 1.1); this interval is termed the voice-onset-time (VOT). In English-speaking adults, chinchillas, and monkeys, the category boundary between /ba/ and /pa/ is around 30 ms (Morse and Snowdon 1975; Kuhl and Miller 1978). This means that for VOT < 30 ms the phoneme is mostly perceived as /ba/, and for VOT > 30 ms as /pa/. In ANFs, neural discharge synchrony (Sinex and McDonald 1989) did *not* provide as much information about VOT as was provided by the best average firing rates (Sinex and McDonald 1988). Many aspects of the patterns of response elicited by VOT syllables in the

central nucleus of the inferior colliculus (ICC; Chen and Sinex 1999) resembled the patterns elicited from ANFs with the same syllables. In the auditory cortex, the neural response is typical of the double onset type; an onset response to the initial noise burst and a second-on response at the start of the vowel (Steinschneider et al. 1982, 1990, 1994; Eggermont 1995a). In other languages, e.g., French and Dutch, the VOT for /ba/ is negative. More will be presented in Chapter 4. We will also see that gap detection is related to forward masking, which in turn reflects perstimulatory adaptation and recovery therefrom (Chapter 3). A relatively simple computational model that relates the major aspects of adaptation, forward masking, gap detection, and temporal modulation transfer functions is presented in Chapter 6.

1.1.1.2 Periodicity representation

Figure 1.1 shows on the right-hand side the periodicity in the vowel /a/, resulting from vocal cord pulsations at a rate of 100–125/s in the male human voice. On the left a kitten vocalization shows a much higher glottal pulse rate (~550/s). Temporal representations of carrier periods exist for most of the auditory pathways alongside rate–place representations but with a decreasing upper frequency limit at more downstream locations. In ANFs, the timing of action potentials is locked to both the period of the carrier frequency—up to 5 kHz (Johnson 1980)—and to the sound envelope, but then only time-locked up to about 2.2 kHz (Joris et al. 2004). In the auditory cortex, the upper limit for unit activity appears to be around 300 Hz but so far has only been demonstrated in the awake monkey's auditory cortex (Steinschneider et al. 1980) and in the ketamine-anesthetized guinea pig thalamus's primary auditory cortex (A1) and its belt areas (Wallace et al. 2000, 2002). The upper limit of phase-locking in the guinea pig encompasses the fundamental frequency of their vocalizations. If the same would hold in cats then one would expect a cortical region with phase-locking well in excess of 500 Hz. This has not been demonstrated to date.

Periodicities in the sound envelope give rise to three dominant percepts: rhythm, roughness, and periodicity pitch. Rhythm, also called fluctuation strength, is perceived for repetition rates below 20 Hz, while the sensation of roughness occurs for rates of 20–300 Hz and is strongest around 70 Hz. Periodicity pitch starting above 30 Hz loses much of its perceptual strength above 3 kHz (Zwicker and Fastl 1990). Thus, there is an overlap of the perception of roughness and pitch, likely due to the co-existence of both temporal (in the ANF firing times) and spectral (place in the cochlea) representations of pitch. Temporal representation of this periodicity in the auditory nervous system manifests itself in two forms; the first is based on the locking of neuronal firings to the period of pure tones or to the fine structure of complex sounds, and this is commonly referred to as phase-locking. It is assumed that the limiting rate of phase-locked responses in ANFs may be related to the upper perceptual pitch boundary. The second form of temporal representation is the locking to the, slower, amplitude modulation (AM) in a complex stimulus; a phenomenon that I will refer to as envelope-locking. Thus, the temporal structure of sound comprises that of the carrier, determining the fine structure or texture of the sound, and that of the stimulus

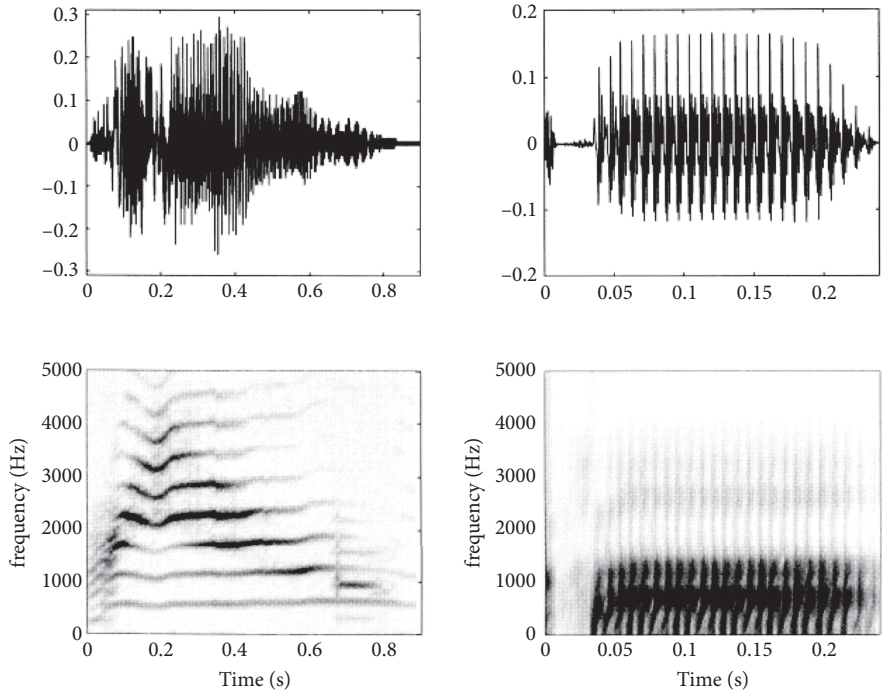

Figure 1.1 Two vocalizations that illustrate similarities and differences in periodicity (contours) and harmonic structure (texture). In the left-hand column, the waveform and spectrogram of a kitten's meow are presented. The duration of this meow is 0.87 s, the average fundamental frequency (F_0) is 550 Hz, and the highest frequency component (not shown) is 5.2 kHz. The two harmonics between 1.5 and 2.5 kHz have the highest intensity. Distinct downward and upward FMs occur simultaneously in all formants between 100 and 200 ms after onset. The meow has also a slow AM. In the right-hand column, the waveform of a /pa/ syllable with a 30 ms VOT and its spectrogram are shown. The periodicity of the vowel and the VOT are evident from the waveform. Note the different timescale. Low-level aspiration noise was present in the period before the onset of voicing. The dominant frequency ranges are F_0 = 125 Hz, F_1 = 700 Hz, and F_2 = 1200 Hz. Because the grey-scale range of the representation is only 30 dB, the third formant at 2600 Hz is only weakly visible. F_0 started at 125 Hz and remained at that value for 100 ms, and dropped from there to 100 Hz at the end of the vowel. F_1 started at 512 Hz and increased in 25 ms to 700 Hz, F_2 started at 1019 Hz and increased in 25 ms to 1200 Hz, and F_3 changed in the same time span from 2153 Hz to 2600 Hz. Reprinted from *Hearing Research*, 157 (1), Jos J. Eggermont, Between sound and perception: reviewing the search for a neural code, pp. 1–42, Copyright (2001), with permission from Elsevier.

envelope that determines the contour of the sound (Eggermont 2001). McDermott and Simoncelli (2011) described the texture of sound as:

> produced by a superposition of many similar acoustic events, such as arise from rain, fire, or a swamp full of insects, and are analogous to the visual textures that have been studied for decades (Julesz, 1962). Textures are a rich and varied set of sounds, and we show here that listeners can readily recognize them. However, unlike the sound of an individual event, such as a footstep, or of

> the complex temporal sequences of speech or music, a texture is defined by properties that remain constant over time. Textures thus possess a simplicity relative to other natural sounds that makes them a useful starting point for studying auditory representation and sound recognition.

Limitations of periodicity-following by the auditory nervous system are typically the result of adaptation, which is the perstimulatory decrease in the response strength and temporal accuracy of the firing of neurons. This is a form of nonstationarity, and the steady state reached after a few 100 ms results in values that are the entries for the temporal modulation transfer functions (tMTF), i.e., "steady-state" synchronization limits to repetitive tone-pips, periodic clicks, or amplitude modulations of noise or high-frequency tones. Since the auditory system is highly non-linear, the simple "linear" description by the tMTFs does not suffice (Forrest and Green 1987) as a characterization. There is more on this in Chapter 5.

1.1.2 Neural coding of sound

Another type of temporal representation of sound is that of a neural code, which does represent sound in a more abstract way compared to the phase-locking that preserves "the relevant time structure of the signal in the cadence of spike discharges evoked by the sound" (Phillips 1993a). Neural coding, however, has been difficult to define (Eggermont 1998b, 2001). Some early proposals are quoted here: ". . . transformed and transmitted signals do not constitute a bonafide neural code unless that information is appropriately acted upon—interpreted—by appropriate parts of the organism" (Perkel and Bullock 1969); ". . . we shall operationally define a true code as a parameter of the signal which actually carries behaviorally usable information" (Uttal 1969); ". . . a neural code at a particular location within the sensory pathway as being the parameter of the afferent discharge in the population that is actually used by the organism in particular sensory behavior" (Mountcastle 1975). More recently, Brugge (1992) stated that

> in order to qualify as a neural code for acoustic information, it must be shown first that the (neural) pattern in question occurs in the auditory system under natural conditions or is evoked by natural stimuli, and second that there exists a sensitive receiver; that is a set of neurons whose activity changes in response to the candidate code it receives.

A more compact statement is: "Codes are the functional organizations that actually utilize a particular set of signs to effect a perceptual discrimination" (Cariani 1995). All cited sources agree that a code relates neural activity to behavior, even if not overt such as in sensory discrimination. Mauk and Buonomano (2004) offered a Morse code as an example in which language is reduced to a temporal code:

> Morse code requires discrimination of continuous streams of sounds and discrimination of the duration, interval, number, and sequence of elements, as well as temporal invariance. The complexity of this analysis provides an example of the sophistication of temporal processing on the timescale of tens to hundreds of ms.

1.1.2.1 Rate coding in the auditory system

The rate-coding hypothesis in its simplest form (Adrian 1928) states that the only important characteristic of a spike train is its mean firing rate. According to this view, both

encoding and decoding are straightforward. The stimulus is encoded by a firing rate proportional to the value of some stimulus parameter, e.g., sound level, and the neuronal response is decoded by counting the spikes. Thus, the set of frequency-tuned neurons that fires represents the sound spectrum and level. Coding by average discharge rate alone necessitates a "labeled line" or "place" coding, because there is no other means within the spike train itself for conveying what kind of signal it is representing (Eggermont 1998b). Only neurons or populations of neurons that have long integration times can interpret rate codes, i.e., can count. Neurons are often modeled as *leaky integrators* with time constants that vary throughout the nervous system. In the periphery, neurons have short time constants (a few milliseconds) and so do most interneurons in the cortex. In contrast, cortical pyramidal cells have time constants of ≥ 10 ms (Mason et al. 1991; Thomson and West 1993). One could therefore ask if rate coding is limited to the cortex, and whether such *place codes* are relevant.

The exclusive use of rate–place codes in the auditory system would face several problems. For instance, periodicity pitch is present only in the temporal characteristics of firing and cannot be conveyed by labeled lines alone. Furthermore, contrast degradation may occur because spontaneous firing and saturation of firing rates limit the dynamic range. It has been shown that a rate–place code for vowel representation largely fails at higher sound levels because of contrast degradation; all firing rates become saturated (Sachs and Young 1979). However, a shift of emphasis from low-threshold high spontaneous firing rate (SFR) fibers to high-threshold low SFR fibers showed a strong potential for rate as a code (Le Prell et al. 1996). How this required shift of emphasis is made obvious to the receiving neurons in the ventral cochlear nucleus is far from clear. In contrast, sound-level coding for pure tones in the auditory nerve (Winslow and Sachs 1988) or midbrain (Eggermont 1989) seems to be based largely on firing rate and to work with small ensembles of neurons (about 10 cells), so one can imagine that intensity coding for a complex sound will require the combination of small neuron pools per frequency into larger pools within a critical band. Critical bands are a measure of the bandwidth of the basic frequency channels in the auditory system and are generally wider than the bandwidth of individual nerve fiber frequency-tuning curves.

1.1.2.2 Temporal coding in the auditory system

To go beyond rate coding, one can examine the statistical relations of the spikes to one another by using inter-spike interval (ISI) distributions for individual neurons or cross-correlation functions for simultaneously recorded spike trains (DeCharms and Zador 2000). Temporal coding relies on coincidence-detecting neurons with short integration times, i.e., shorter than the modal ISI of a neuron. For cortical neurons the modal ISI may be of the order of a few milliseconds, whereas the integration time of pyramidal neurons is of the order of 10 ms. This suggests that coincidence detection, and thus temporal coding, is therefore limited to the sensory periphery, auditory brainstem, and to the interneurons in the cortex, which have much shorter integration times. Pyramidal cells in cortex may be detecting coincidences between direct thalamic inputs and indirect (delayed) inputs from

local interneurons, association fibers, and commissural fibers (König et al. 1996). An array of coincidence elements can compute correlations or convolutions, since these only differ by a sign change of the time in one of the spike trains. A population of coincidence elements embedded in a system of relative delays can compute global auto- and cross-correlation functions, which may form the basis for complex feature extraction. True coincidence detection is, however, not optimal; an optimal multiplier neuron encodes a signal formed by smoothly weighting all of the near misses of coincidences between the spike trains (Bialek and Rieke 1992). If cortical neurons behave as coincidence detectors, then the timing of spikes can propagate through the cortex with great fidelity to convey information and to synchronize other neurons (Abeles 1982). We will discuss this further in Chapter 7.

There are, however, exceptions to the suggestion that only synchrony, i.e., coincidence detection, or mechanisms relying upon synchrony can be used in speech coding. For instance, speech features such as stop-bursts and frication at the level of the auditory nerve may be better encoded in terms of a rate–place profile (Delgutte and Kiang 1984). Because in the primary auditory cortex neither the fundamental nor the formants can be represented in temporal fashion, a transformation from temporal to place code between auditory nerve and cortex is needed to solve this problem. In the auditory system action potentials in different neurons are generally produced synchronously with periodicities in the sound and thus with each other. Timing and synchrony in neural populations thus contain information about these periodicities.

1.1.2.3 Envelope coding in the auditory system

Auditory thalamic and cortical neurons can exhibit extremely reliable temporal patterns locked to stimulus envelopes across several repetitions of a natural sound. These patterns are underlying a temporal coding scheme that is far more efficient than using spike counts for certain stimuli (Huetz et al. 2009). If one assumes that neurons are implementing a rate-coding scheme for simple spectro-temporal features of the stimulus, or, in other words, act as spectro-temporal filters of the stimuli, the evoked spike timing should be reliable across trials and should differ for different stimuli (Figure 1.2).

Gehr et al. (2000) investigated the representation of natural and morphed cat vocalizations in single-unit activity of the ketamine-anesthetized cat's primary auditory cortex. About 40% of the neurons showed time-locked responses to major peaks in the vocalization envelope, while 60% only responded at the onset. Changing the temporal properties of the stimuli, as in the forward vs. time-reversed and expanded vs. compressed meows, did not dramatically change the firing rate in comparison to the natural meow (Figure 1.3). It was obvious that acoustical stimulation influenced the neurons, because the firing rates were significantly higher than those found under spontaneous conditions. A special temporal coding related to the waveforms of the stimuli seemed possible for the set of peak-tracking units. This set of units also carried on average 25–50% more information about the stimulus in its firing rate. The representation of the vocalizations seemed to take place in a synchronized and distributed fashion, because of the significantly enhanced correlations between different recording sites, with separations up to 2 mm, during stimulation.

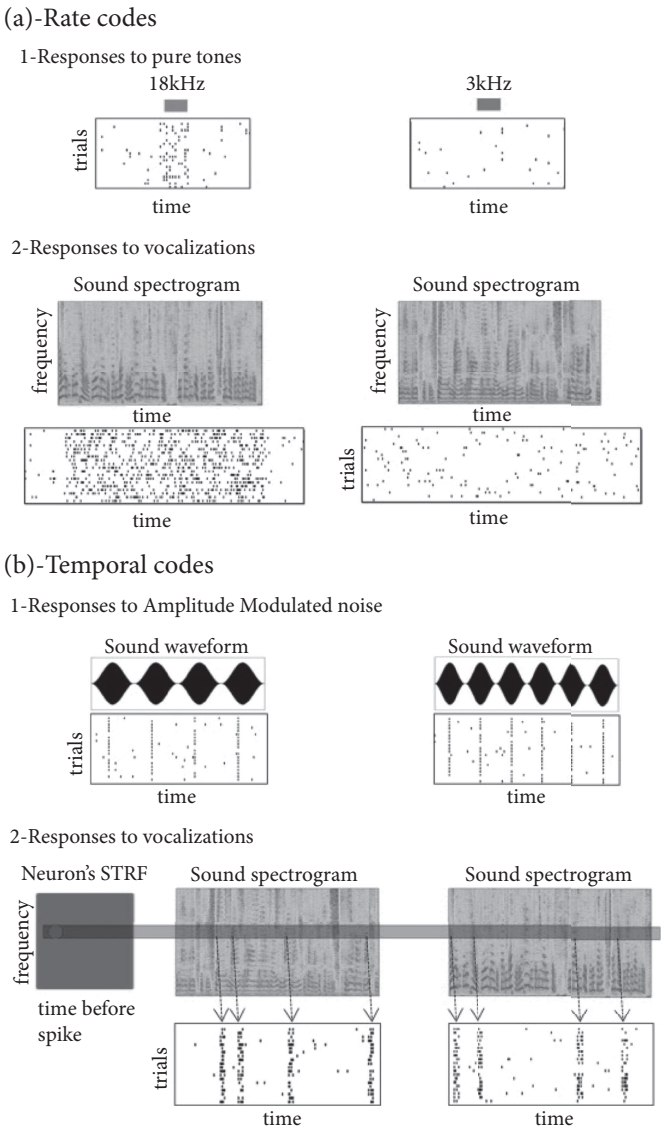

Figure 1.2 Outlines of the rate (A) and temporal code (B) hypotheses. Each panel displays the responses of an artificial neuron over 20 repetitions of the stimulus represented. (A) Rate code hypothesis applied to artificial stimuli (1) and natural sounds (2). The artificial neuron increases its firing rate in a Poisson-like way for some stimuli (on the left side) but not others (right). In this case, the only variable allowing discrimination between stimuli is the overall firing rate.
(1) Responses of an artificial neuron to SAM noise of different frequencies. The artificial neuron emits a spike in a reliable manner at each phase of the modulated noise, and therefore shows reliable temporal patterns. (2) Responses of an artificial neuron to two vocalizations. The neuron's STRF (on the left panel) shows a response for a specific frequency range. If this neuron is tested with natural stimuli (middle and right panels), it will emit reliable temporal patterns that can be discriminated between stimuli. Reprinted from *Hearing Research*, 271 (1–2), Chloé Huetz, Boris Gourévitch, and Jean-Marc Edeline, Neural codes in the thalamocortical auditory system: From artificial stimuli to communication sounds, pp. 147–58, Copyright (2011), with permission from Elsevier.

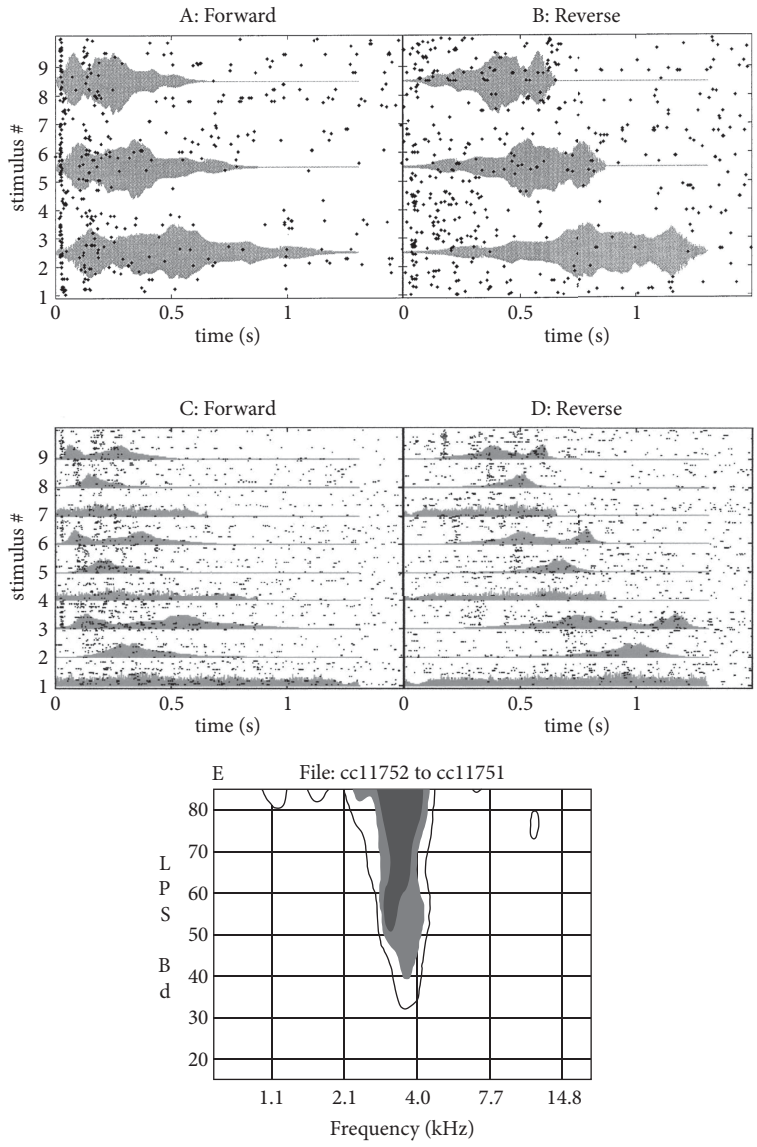

Figure 1.3 Dot displays of the responses to (A) the forward meow and (B) the time-reversed meows at one MU recording site (top panels). Time (x-axis) is specified in seconds. The different stimuli are numbered from 1 to 9. Stimuli 1–3 are the expanded meows with lowered (stimulus 1), unaltered (stimulus 2), and increased (stimulus 3) frequency components. Stimuli 4–6 are unaltered in time and stimuli 7–9 are time-compressed. The waveform envelopes of stimuli in the top two panels have compressed (stimulus 8), normal (stimulus 5), and expanded (stimulus 2) envelope duration and are shown in a gray shade. The panels C and D in the second row show stimuli that are filtered through the tuning curve of the neuron (E) and the positive parts of the waveform are shown superimposed on the dot rasters. Note that this represents a different recording from that in the top row. The envelope waveforms are scaled individually to the maximum. The frequency response area shown in E shows the tuning curve contour at 25% of the maximum firing rate, and the area with > 50% of the peak firing rate (black). Reprinted from *Hearing Research*, 150 (1–2), Daniel D Gehr, Hisashi Komiya, and Jos J Eggermont, Neuronal responses of cat primary auditory cortex to natural and altered species-specific calls, pp. 27–42, Copyright (2000), with permission from Elsevier.

This increase in correlation was, however, entirely due to those for the peak-tracking neurons, suggesting that if synchronized ensemble formation takes place it is restricted to this subpopulation. The sum of the responses to the low- and high-frequency part of the meow, with the boundary at 2.5 kHz, was larger than the neuronal response to the natural meow itself, suggesting that strong lateral inhibition is shaping the response to the natural meow. The study suggested that for cat meows the relevant aspects are the excitatory and inhibitory frequency tuning and the tMTF of the neuron, together with the synchronization of the activity of a subset of peak-tracking neurons.

The potential effect of the frequency-tuning properties of the units on the evoked firings is illustrated in Figure 1.3 (CD). The set of meow envelopes shown in Figure 1.3 (AB) (forward and time-reversed morphed meow sequences) were passed through a band-pass filter mimicking the tuning curve of an MU recording site narrowly tuned to 3.7 kHz (Figure 1.3 (E)). The one-sided envelopes of the filtered meows are shown together with the MU dot rasters. The low-frequency content stimuli (Nos 1, 4, and 7) only present an unmodulated envelope, whereas the natural and high-frequency content vocalizations show distinctly different envelopes after filtering. The neural responses are more clearly related to the envelopes of the filtered vocalizations than those of the unfiltered stimuli (as in Figure 1.3 (AB)). Once the stimulus reaches a certain threshold, the units show an excitatory response. In general, filtering of the morphed meow sounds by the MU frequency-tuning curve of the recording site clearly predicted the range of onset latency values that was observed across the population (Gehr et al. 2000).

Gourévitch and Eggermont (2007a) subsequently showed the spatial and temporal representation of cat vocalizations, natural and altered with respect to carrier and envelope, as well as time-reversed (using the same stimuli as in Gehr et al. 2000), in the auditory cortex. Multiunit activity recorded in AI of ketamine-anesthetized cats again occurred mainly at onset and at subsequent major peaks of the vocalization envelope, and was significantly inhibited during the stationary course of the stimuli. The first 200 ms of processing appeared crucial for discrimination of a vocalization in AI. Discrimination of the neural responses to different alterations of vocalizations could be based on either firing rate, type of temporal response, or neural synchrony, suggesting that all these are likely to be used simultaneously in the processing of natural and altered conspecific vocalizations. Spontaneous synchrony is normally weak between distant recording sites in AI (Eggermont 1992). Vocalization processing induced a general increase of synchrony between neurons compared with the spontaneous firing condition. The increase in synchrony was not caused by increased firing rate because the correction procedure used for the cross-correlation coefficient eliminated such effects. The increase in synchrony remained constant over stimulus time. We showed that post-stimulus-time histogram (PSTH) type (temporal information), firing rate, and neural synchrony allowed discrimination of the features of altered and all time-reversed vocalizations. All these parameters and perhaps some others might be involved simultaneously in the processing and decoding of such complex sounds as vocalizations. Huetz et al. (2009) applied information theory-based analyses to single-unit spike trains collected in the auditory cortex and auditory thalamus of anesthetized guinea pigs as well as in the auditory cortex of awake guinea pigs during presentation of four

conspecific vocalizations. Few thalamic and cortical cells (<10%) displayed a firing rate preference for the natural version of these vocalizations. In contrast, when the information transmitted by the spike trains was quantified with a temporal precision of 10–50 ms, many cells (>75%) displayed a significant amount of information (i.e., >2SD above chance levels), especially in the awake condition. Based on temporal discharge patterns, even cells that were only weakly responsive to vocalizations displayed a significant level of information. These findings also emphasized the importance of temporal discharge patterns as a coding mechanism for natural communication sounds, particularly in awake animals.

1.2 Pitch theories: time and/or frequency representation

Historically, theories of hearing have always been theories of pitch perception. For a pure tone with frequency below a few thousand Hz, the pitch was considered to reflect either the place of activation on the basilar membrane (Helmholtz 1863) or the periodicity of the firings of the auditory nerve that are phase-locked with the tone period (Wundt 1880). For pure tones—with frequencies below 3 kHz—there is really no way to favor one theory above the other. However, most sounds in nature do not consist of pure tones, but are rather a mix of harmonic frequencies forming a complex tone as produced by, e.g., pianos, string and wind instruments, or animal vocalizations (Figure 1.1). Early on, Seebeck (1841) had realized that the mechanisms underlying the pitch of complex tones might not be straightforward. The pitch corresponding to the period of acoustic pulses of a siren disk with equidistant holes is heard much stronger than the individual harmonics. Thus, Seebeck (1843) concluded that the physical presence of a sinusoidal component with frequency f was not essential for hearing a pitch corresponding to this frequency. Earlier in the same year Ohm (1843) had applied the Fourier principle to the case of the frequency analysis of the ear, and proposed that a sound component given by $\sin(2\pi ft + \varphi)$ is required to hear a pitch corresponding to the frequency f. Ohm's acoustic law was endorsed by Helmholtz (1863). He was the first to suggest a mechanism by which the ear arrives at a frequency analysis, the resonance–place theory. Postulating resonance as the mode of analysis, Helmholtz thought that in response to a particular frequency, a well-localized displacement maximum of the basilar membrane occurred. This maximum would be shifting with the frequency of the sound. On the basis of psychoacoustic evidence, Helmholtz estimated the sharpness of the resonators as having a bandwidth of 4% of the resonance frequency, and thus being proportional to frequency. Helmholtz also assumed that the local vibrations of the basilar membrane in the cochlea gave rise to activity of the corresponding ANFs. Helmholtz hypothesized that a specific pitch corresponds to each of the nerve fibers that make contact with (the hair cells in) the cochlea. Helmholtz's theory correlated frequency with place, and place with pitch, so it assumed a frequency-pitch correspondence (Plomp 1968).

In 1928 and 1929 the classic papers by von Békésy (appearing in German in the *Physikalische Zeitschrift* and translated by Wever in von Békésy 1960) on the biophysics of the inner ear, the traveling wave, and the frequency-specific maximum displacement on the basilar membrane validated the assumptions of Helmholtz.

Back to pitch theories, Wundt (1880) proposed an alternative explanation of pitch, interestingly based on a finding by Helmholtz that was unrelated to the auditory system. Helmholtz (1868) had found that stimulating muscles via their nerves with periodic electric pulses gave rise to synchronous mechanical vibrations of the muscle. Up to stimulation with about 240 pulses-per-second, these vibrations manifested themselves by a clear tone emitted from the muscle. In the second edition of his work, "*Grundzuge der Physiologischen Psychologie*", Wundt criticized Helmholtz's hypothesis that cochlear place is correlated with pitch. While accepting the hypothesis that different locations along the basilar membrane are tuned to different frequencies, Wundt proposed an alternative for Helmholtz's place–pitch hypothesis, namely that tones give rise to synchronous nerve impulses whose periodicity determines pitch. The pitch perception theory promoted by Wundt assumed a periodicity–pitch correspondence. Although Wundt's theory could explain some observations with complex sounds much better than the resonance theory, its influence was much less than that of Helmholtz (Plomp 1968). Wundt's theory was resurrected by his student Wever (Wever and Bray 1930) and evolved in the already-mentioned "Volley Theory" of hearing (Wever 1949). Independently, Rutherford (1886) had also favored the periodicity theory:

> The theory which the lecturer had arrived at, and which he published that night for the first time, might be termed the telephone theory of the sense of hearing—the theory that the cochlea does not act on the principle of sympathetic vibration, but that the hairs of all its auditory cells vibrate to every tone just as the drum of the ear does; that there is no analysis of complex vibrations in the cochlea or elsewhere in the peripheral mechanism of the ear; that the hair cells transform sound vibrations into nerve-vibrations similar in frequency and amplitude to the sound vibrations; that simple and complex vibrations of nerve energy arrive in the sensory cells of the brain, and there produce, not sound again of course, but the sensations of sound, the nature of which depends not upon the stimulation of different sensory cells, but on the frequency, amplitude, and form of the vibrations coming into the cells, probably through all the fibres of the auditory nerve.

Close reading of this quote suggests that, according to Rutherford, one hair cell and one ANF would be sufficient to represent the entire frequency range.

Modern temporal pitch models combine interspike intervals from single auditory neurons in all frequency regions to produce "population interval distributions," i.e., global distributions of intervals in neural populations. These models can be based on either first-order intervals, i.e., between successive spikes, or all-order intervals, i.e., between successive and non-successive spikes. All-order intervals constitute the autocorrelation function. Licklider (1951) described it as:

> The essence of the duplex theory of pitch perception is that the auditory system employs both frequency analysis and autocorrelational analysis. The frequency analysis is performed by the cochlea, the autocorrelational analysis by the neural part of the system.

These models are based on a "predominant interval" hypothesis, and the pitch heard corresponds to the most frequent interval in a population interval distribution. All of these purely temporal models successfully predict pitch for a wide range of complex stimuli, reinforcing the plausibility that the predominant interval hypothesis might hold at the

level of the auditory nerve (Cariani and Delgutte 1996). This topic will be further explored in Chapter 8.

1.3 Spectro-temporal or texture–contour coding of sound?

It is currently still up in the air whether the best neural coding strategy is based on firing rate or on firing times and neural synchrony. Based on what we discussed, the type of code has to change going from auditory nerve to cortex. The assumption of sound coding based on spike timing led to procedures for sound reconstruction on the basis of spike trains. It all started with changes in the philosophy behind the way the auditory system is experimentally assessed (Eggermont et al. 1983):

> We will call this [old] approach the *experimenter-centred* approach; the experimentally observed variable is the occurrence of action potentials (spikes), commonly expressed as a firing rate. One may also use criteria as just detectable changes in firing rate, e.g. in the construction of frequency-tuning curves, and more recently also changes in synchrony between the occurrence of the spikes and the stimulus. The stimuli used in this approach are the familiar ones such as clicks, tone- and noise-bursts or continuous tones of various frequencies. . . . One calculates the occurrence of spike activity as a function of time, τ, *after* the onset of stimulus presentation. The experimenter-centred approach is therefore a method of *forward correlation*. The second [new] approach, pioneered by De Boer (1967, 1968), . . . may be called a *subject-centred* approach. This way of thinking considers the occurrence of an action potential (event) as a sign or as an indication that something particular happened with the stimulus preceding that action potential. Each action potential is considered as signalling a stimulus that was of interest to the neuron and may be even to the animal. For this type of approach the stimulus preferably is diverse in nature, and the pioneering papers reported results based on Gaussian wide-band noise as a stimulus. By correlating the events with the stimulus in a given interval prior to the event one obtains an estimate of the average stimulus that caused the spikes. This correlation procedure where one looks at the average stimulus as a function of time, τ, prior to the spike is therefore a method of backward- or *reverse-correlation* (de Boer 1968).

This is the domain of "reverse correlation" either in the time domain (de Boer and Kuyper 1968) or in the spectro-temporal domain (Aertsen and Johannesma 1981). In the latter case this results in the spectro-temporal receptive field (STRF). The recent surge in the global use of STRFs (Depireux et al. 2001; Depireux and Elhilali 2013) after its first localized use in the early 1980s (Eggermont et al. 1983) to quantify neural response properties in the auditory midbrain follows from these early ideas. Population activity is per definition based on neural synchrony (Tomita and Eggermont 2005). The role of population activity vs. individual neuron activity may change along the auditory pathway. Codes that go beyond the auditory system may involve brain rhythms, which are temporal and rely on population activity based on neural synchrony.

Another dual representation of sound can be made by the already-mentioned distinction of sound texture and sound contour (Eggermont 2001):

> Some of the auditory features, for which we reviewed the neural representation at various stations along the auditory pathway, include common onset and offset of sound, common rates of AM and FM, harmonicity and common spatial origin. It is useful for the subsequent synthesis to