SPRINGER HANDBOOK OF AUDITORY RESEARCH

Series Editors: Richard R. Fay and Arthur N. Popper

David Poeppel
Tobias Overath
Arthur N. Popper
Richard R. Fay
Editors

The Human Auditory Cortex

Springer

Springer Handbook of Auditory Research

For further volumes:
http://www.springer.com/series/2506

David Poeppel • Tobias Overath
Arthur N. Popper • Richard R. Fay
Editors

The Human Auditory Cortex

Springer

Editors
David Poeppel
Department of Psychology
New York University
New York, NY 10003, USA

Tobias Overath
Department of Psychology
New York University
New York, NY 10003, USA

Arthur N. Popper
Department of Biology
University of Maryland
College Park, MD 20742, USA

Richard R. Fay
Marine Biological Laboratory
Woods Hole, MA 02543, USA

ISSN 0947-2657
ISBN 978-1-4614-2313-3 e-ISBN 978-1-4614-2314-0
DOI 10.1007/978-1-4614-2314-0
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2012933577

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

The editors dedicate this book to Sheila E. Blumstein, the Albert D. Mead Professor of Cognitive, Linguistic and Psychological Sciences at Brown University. Sheila has made a number of key contributions to our understanding of speech recognition, word recognition, and aphasia. No colleague has been as effective in using (and bridging) auditory psychophysics, psycholinguistics, data from aphasic patient populations, and functional brain imaging. Her research has motivated many of our experiments and theories, and her scientific citizenship makes her a model for many of us in the field.

Series Preface

The Springer Handbook of Auditory Research presents a series of comprehensive and synthetic reviews of the fundamental topics in modern auditory research. The volumes are aimed at all individuals with interests in hearing research including advanced graduate students, post-doctoral researchers, and clinical investigators. The volumes are intended to introduce new investigators to important aspects of hearing science and to help established investigators to better understand the fundamental theories and data in fields of hearing that they may not normally follow closely.

Each volume presents a particular topic comprehensively, and each serves as a synthetic overview and guide to the literature. As such, the chapters present neither exhaustive data reviews nor original research that has not yet appeared in peer-reviewed journals. The volumes focus on topics that have developed a solid data and conceptual foundation rather than on those for which a literature is only beginning to develop. New research areas will be covered on a timely basis in the series as they begin to mature.

Each volume in the series consists of a few substantial chapters on a particular topic. In some cases, the topics will be ones of traditional interest for which there is a substantial body of data and theory, such as auditory neuroanatomy (Vol. 1) and neurophysiology (Vol. 2). Other volumes in the series deal with topics that have begun to mature more recently, such as development, plasticity, and computational models of neural processing. In many cases, the series editors are joined by a co-editor having special expertise in the topic of the volume.

Woods Hole, MA, USA — Richard R. Fay
College Park, MD, USA — Arthur N. Popper

Volume Preface

This volume brings the Springer Handbook of Auditory Research series to its first detailed examination of auditory cortex, with a strong emphasis on auditory processing in humans. Chapters are grouped into two sections or themes (methods and content areas), although, as seen in reading the chapters, many actually present material that covers the two general themes of the volume. Chapters 2 to 6 in Section I explain the main techniques currently available to study the human brain, with a specific focus on their use in investigating auditory processing. Chapters 7 to 13 in Section II cover different aspects of auditory perception and cognition.

In Chapter 2, Clarke and Morosan describe the anatomy of the human auditory cortex and introduce the nomenclature of the different subareas in the human auditory cortex, in a historical context. In Chapter 3, Howard, Nourski, and Brugge introduce a type of data that are extremely rare: intracranial recordings from patients undergoing neurosurgical procedures or presurgical evaluation.

Chapters 4 through 6 provide descriptions of the most typical methodologies available to study human auditory perception and cognition. Electroencephalography (EEG) is the topic of Chapter 4 by Alain and Winkler. The other noninvasive electrophysiological technique that is increasingly widespread is magnetoencephalography (MEG), outlined in Chapter 5 by Nagarajan, Gabriel, and Herman. In Chapter 6, Talavage, Johnsrude, and Gonzalez turn to functional magnetic resonance imaging (fMRI), an approach that provides excellent spatial resolution.

In Chapter 7, Hall and Barker discuss the neural processing of basic perceptual attributes and familiarize the reader with elementary problems such as the encoding of pure tones, pitch, and loudness. Following this, the concept of auditory objects or auditory streams is considered in Chapter 8 by Griffiths, Micheyl, and Overath.

One of the major naturalistic tasks for the auditory system, speech perception, is the topic of Chapter 9 by Giraud and Poeppel. The other auditory domain receiving a great deal of attention and generating a fascinating body of data concerns the cortical foundations of processing music, as discussed in Chapter 10 by Zatorre and Zarate. This is followed by a consideration of the multisensory role of the human auditory cortex by van Wassenhove and Schroeder in Chapter 11. In Chapter

12, Hickok and Saberi discuss the variety of function of the planum temporale, an area of the human cortex considered integral to processing many aspects of complex sounds.

The volume concludes with a consideration by Cariani and Micheyl in Chapter 13 of what is known with respect to computational models that link the data from the multiple methodologies in human and animal models in an explicit way.

A number of recent volumes in the Springer Handbook of Auditory Research series complement, and are complemented by, this volume on *Human Auditory Cortex*. Computation in the auditory system is considered at length in *Computational Models of the Auditory System* (Vol. 35, edited by Meddis, Lopez-Pevada, Fay, and Popper, 2010), whereas perception of music is the topic of *Music Perception* (Vol. 36, edited by Riess-Jones, Fay, and Popper, 2010). Human perception and sound analysis are considered most recently in *Loudness* (Vol. 37, edited by Florentine, Popper, and Fay, 2011), *Auditory Perception of Sound Sources* (Vol. 29, edited by Yost, Popper, and Fay, 2008), and *Pitch: Neural Coding and Perception* (Vol. 21, edited by Plack, Oxenham, Fay, and Popper, 2005).

New York, NY, USA	David Poeppel
New York, NY, USA	Tobias Overath
College Park, MD, USA	Arthur N. Popper
Woods Hole, MA, USA	Richard R. Fay

Contents

Contributors

Claude Alain Rotman Research Institute, Baycrest Centre for Geriatric Care, 3560 Bathurst Street, Toronto, ON M6A 2E1, Canada

Department of Psychology, University of Toronto, Ontario M8V 2S4, Canada

Daphne Barker School of Psychological Sciences, University of Manchester, Manchester M13 9PL, UK

John F. Brugge Department of Neurosurgery, University of Iowa, 200 Hawkins Dr. 1624 JCP, Iowa City, IA 52242, USA

Peter Cariani Department of Otology and Laryngology, Harvard Medical School, 629 Watertown Street, Newton, MA 02460, USA

Javier Gonzalez-Castillo Section on Functional Imaging Methods, Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, Bethesda, MD 20892, USA

Stephanie Clarke Service de Neuropsychologie et de Neuroréhabilitation, CHUV, 1011 Lausanne, Switzerland

Rodney A. Gabriel Department of Radiology and Biomedical Imaging, University of California, San Francisco, 513 Parnassus Avenue, S362, San Francisco, CA 94143, USA

Anne-Lise Giraud Inserm U960, Département d'Etudes Cognitives, Ecole Normale Supérieure, 29 rue d'Ulm, 75005 Paris, France

Timothy D. Griffiths Institute of Neuroscience, Newcastle University Medical School, Framlington Place, Newcastle upon Tyne, NE2 4HH, UK

Deborah Hall NIHR National Biomedical Research Unit in Hearing, Nottingham NG1 5DU, UK

Alexander Herman Department of Radiology and Biomedical Imaging, University of California, San Francisco, 513 Parnassus Avenue, S362, San Francisco, CA 94143, USA

Gregory Hickok Department of Cognitive Sciences, University of California, Irvine, CA 92697, USA

Matthew A. Howard III Department of Neurosurgery, University of Iowa, 200 Hawkins Drive, 1823 JPP, Iowa City, IA 52242, USA

Ingrid S. Johnsrude Department of Psychology, Queen's University, Kingston, ON K7L 3N6, Canada

Christophe Micheyl Auditory Perception and Cognition Laboratory, Department of Psychology, University of Minnesota, N628 Elliot Hall, 75 East River Road, Minneapolis, MN 55455, USA

Patricia Morosan Institute of Neurosciences and Medicine (INM-1), Research Centre Jülich, 52425 Jülich, Germany

Srikantan Nagarajan Department of Radiology and Biomedical Imaging, University of California, San Francisco, 513 Parnassus Avenue, S362, San Francisco, CA 94143, USA

Kirill V. Nourski Department of Neurosurgery, University of Iowa, 200 Hawkins Drive, 1815 JCP, Iowa City, IA 52242, USA

Tobias Overath Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, USA

David Poeppel Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, USA

Kourosh Saberi Department of Cognitive Sciences, University of California, Irvine, CA 92697, USA

Charles E. Schroeder Professor of Psychiatry, Cognitive Neuroscience and Schizophrenia Program, Nathan S. Kline Institute for Psychiatric Research, 140 Old Orangeburg Road, Orangeburg, NY 10962, USA

Thomas M. Talavage School of Electrical & Computer Engineering, Purdue University, West Lafayette, IN 47907, USA

Virginie van Wassenhove CEA DSV.^{I2}BM.NeuroSpin, Cognitive Neuroimaging Unit (INSERM U992), Bât 145 - Point Courrier 156, Gif s/Yvette F-91191, France

István Winkler Institute for Psychology, Hungarian Academy of Sciences, P.O. Box 398, H-1394 Budapest, Hungary

Institute of Psychology, University of Szeged, 6722 Szeged, Petőfi S. sgt. 30-34, Hungary

Jean Mary Zarate Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, USA

Robert J. Zatorre Cognitive Neuroscience Unit, Montréal Neurological Institute, McGill University, 3801 reu University, Montréal, Québec H3A 2B4, Canada

Chapter 1
Introduction: Why Human Auditory Cortex?

David Poeppel and Tobias Overath

This volume concentrates on current approaches to understanding the human auditory cortex.

Why auditory? The reasons why one would, could, and should study auditory processing ought to require little comment or motivation. That being said—and given the critical importance of hearing and speech for human communication and welfare—one might wonder why hearing research, in general, has remained the less popular stepchild and ugly duckling in the context of sensory neuroscience, in particular vision. (This question is discussed in more detail later.).

Why human? Our knowledge of nonhuman hearing is increasingly well developed, and there now exist excellent recent overviews of auditory processing at various levels of analysis, ranging from anatomy to neurophysiology to computational modeling of subtle hearing phenomena (e.g., Oertel et al., 2002; Manley et al., 2008; Meddis et al., 2010; Moore 2010; Winer & Schreiner 2010; Schnupp et al., 2011). The very existence of the extensive and thorough *Springer Handbook of Auditory Research* is a testament to the fact that this research arena is perceived as a growth area at the cutting edge of the behavioral and brain sciences. Yet, for human auditory processing, although our behavioral/psychophysical approaches are sophisticated, our *knowledge of the neural basis is still quite rudimentary*. A detailed focus on the human auditory system seems timely and necessary.

Why cortex? It goes without saying that the contributions of subcortical structures to virtually all aspects of human auditory perception are immense. The afferent auditory pathway is highly complex and richly structured; it is sometimes argued, for example, that the inferior colliculus constitutes a better comparison to primary visual cortex than primary auditory cortex itself (King & Nelken, 2009). Moreover,

D. Poeppel (✉) • T. Overath
Department of Psychology, New York University, 6 Washington Place,
New York, NY 10003, USA
e-mail: david.poeppel@nyu.edu; t.overath@nyu.edu

D. Poeppel et al. (eds.), *The Human Auditory Cortex*, Springer Handbook
of Auditory Research 43, DOI 10.1007/978-1-4614-2314-0_1,

even higher-order tasks, including speech perception, are modulated by subcortical circuitry. Nevertheless, it is *cortical structures that lie at the basis of auditory perception and cognition*. To restrict the scope of coverage, and fully acknowledging the importance of other cerebral regions, the human auditory cortex is the target of this inquiry.

The study of the visual system—across species, levels of visual processing, and approaches—continues to be a dominant focus in the neurosciences. The amount of resources dedicated to vision is easy to understand: humans are highly visual creatures (if the criterion is the proportion of cerebral real estate allocated to visual processing in one form or another); animal models of human vision are remarkably successful, allowing for detailed mechanistic characterizations of various aspects of visual perception across species, including humans; and the experimental "management" of visual materials has made the research tractable for a long time, ranging from the simple holding up of an image to tachistoscopic presentation to sophisticated computer graphics.

Recently, more researchers are turning their attention to auditory processing. By analogy to vision, three trends are worth mentioning. First, investigators now acknowledge that humans are also highly auditory creatures, if the criteria are (1) the amount of cerebral territory implicated in hearing as well as (2) what humans are willing to pay for to be entertained, for example, iTunes®; second, animal models are highlighting important similarities but also showing some key differences between the human auditory system and well-studied nonhuman preparations, notably with respect to complex sound processing, speech, and music; and finally, new technologies have made manipulating auditory materials much easier—and many auditory studies doable to begin with. (Recall that crafting and editing digital audio files with such speed and ease is a rather recent development, in terms of the history of the work.) Moreover, and crucially, the development of new noninvasive recording and imaging techniques to study the human brain has opened up entirely new possibilities for investigating human hearing; the impact of these techniques on research in (human) auditory processing cannot be overstated. One key aspect of this volume is to highlight the existing techniques and illustrate how they are used to study various aspects of human auditory perception.

The book's chapters are roughly organized in two sections, *methodologies* and *content areas*. The chapters in the first section explain the techniques currently available to study the human brain, with a specific focus—and numerous examples—on auditory processing. The coverage is necessarily selective; for example, the volume does not cover recent experimental work using transcranial magnetic stimulation (TMS), transcranial direct current stimulation (tDCS), and near-infrared spectroscopy (NIRS). The chapters were designed to cover the areas of experimental inquiry in which a fairly extensive and robust body of results exists on human auditory cortical structure and function. It stands to reason that newer recording techniques will make major discoveries, but it seems prudent to restrict the focus on methodologies with a significant track record.

In the second group of chapters, different aspects of auditory perception and cognition are discussed, that is, the focus of each chapter lies on a specific content area (e.g., auditory objects, speech, music) or an aspect of auditory processing that cuts across domains (e.g., multisensory perception, perception–action interaction, etc.).

The coverage of the book is as follows. In Chapter 2, Clarke and Morosan describe the neurobiological infrastructure that lies at the very foundation of the entire research area: the anatomy of the human auditory cortex. The data have derived mainly from cytoarchitectonic analyses of postmortem brains, but now also incorporate recent insights from other anatomic approaches. The chapter also introduces, in a historical context, the nomenclature of the different subareas in human auditory cortex. The quip that "anatomy is destiny" is attributed to Sigmund Freud—and, to our knowledge, he was not referring to the structure and function of the auditory system. It is clear, however, that human auditory research needs to be much more granular about the computational contribution each putative cortical region makes. The success and destiny of the research program is indeed predicated on whether it is possible to forge detailed linking hypotheses between anatomic structures and computational subroutines.

In Chapter 3, Howard, Nourski, and Brugge introduce a type of data that most researchers rarely have access to: intracranial recordings from patients undergoing neurosurgical procedures or presurgical evaluation. Although such direct invasive neurophysiological recordings are much harder to come by—and are associated with the typical limitations of the clinical situation—these new data sets are gaining currency and provide a window onto auditory function that allows us to make important connections between noninvasive imaging and the neurophysiological data obtained in animal studies. The chapter introduces the nuts and bolts of how auditory research is done in this context.

Chapters 4, 5, and 6 provide descriptions of the most typical methodologies currently in use to study human auditory perception and cognition. Electroencephalography (EEG), the topic of Chapter 4 by Alain and Winkler, has been available for cognitive neuroscience research since the 1930s. Although its popularity has waxed and waned over the years, it is fair to say that EEG data are now richly appreciated in auditory research and have provided the most data (and perhaps insight) about auditory function, in particular with respect to the temporal properties of perception. The other noninvasive electrophysiological technique that is currently growing in use is magnetoencephalography (MEG), a cousin of EEG. The technique is outlined by Nagarajan, Gabriel, and Herman in Chapter 5, where a number of examples illustrate how MEG can be used to investigate aspects of auditory processing that can be more challenging than with other electrophysiological approaches. The balance between temporal and spatial resolution afforded by MEG makes the technique well suited to investigate research questions in which localization of function plays a role that cannot be addressed effectively with, say, EEG.

In Chapter 6, Talavage, Gonzalez, and Johnsrude turn to the hemodynamic recording approaches, principally functional magnetic resonance imaging (fMRI). The fantastic spatial resolving power of this imaging technique is now well known. Because of some of the quirks of this technique—it is quite loud to be inside the scanner as a participant, and the response that is quantified develops over several seconds (hemodynamics are slow relative to electricity)—the utility of fMRI in auditory research was initially somewhat limited. However, technical innovations in the last few years have rendered this tool an excellent window through which to view human auditory cortex as it processes signals ranging from single tones to extended narratives. The chapter provides a thorough description of the underlying neurophysiology, design, and analysis of fMRI data.

Chapters 7 to 13 form the second part of the book, now with less explicit emphasis on the particular methods, but rather a concentration on specific perceptual and conceptual challenges that the human auditory system faces. In Chapter 7, Hall and Barker discuss the basic acoustic constituents that comprise the auditory environment. The chapter familiarizes the reader with elementary problems such as the encoding of pure tones, the representation of sounds with larger bandwidths and more complex spectral structure, the analysis of pitch, and the effect of loudness. The perceptual attributes described and discussed in this chapter form the basis of representations of an intermediate complexity, lying between encoding at the auditory periphery, on the one hand, and the perceptual interpretation as high-level objects, including speech or music, on the other. The level of representation that is the centerpiece of this chapter might be considered the “perceptual primitives” of auditory cognition; these are the attributes that humans can independently assign to an auditory stimulus, regardless of its category membership. For example, whether humans are characterizing an environmental stimulus, a musical motif, or a spoken word, we can talk about the pitch, the spatial position, or the loudness of the signal.

The concept of what auditory objects or auditory streams actually are—already raised in the description of EEG (Chapter 4)—is tackled more directly in Chapter 8 by Griffiths, Micheyl, and Overath. The question of what constitutes an auditory object has proven to be remarkably controversial. In part, this difficulty may stem because this concept derives by and large from vision research and may not have a one-to-one transfer function to audition (e.g., the temporal dimension is arguably more important in audition). To generalize the notion of object, other dynamic features of an object must now come to fore, enriching the discussion of what is considered to be an elementary representation in a sense that satisfies both visual and auditory theories. The chapter describes the empirical research that has contributed to clarifying the problems, notably work on streaming, grouping, and sequencing. Both hemodynamic and electrophysiological studies are described that aim to elucidate this complex notion.

One of the major naturalistic tasks for the auditory system, speech perception, is the theme of Chapter 9, by Giraud and Poeppel. There, a decidedly neurophysiological perspective is provided, focusing especially on the perceptual analysis of connected speech—and not on how individual vowels, syllables, or words are analyzed. Whereas there exists a fairly large literature on the neurobiological activity associated with perceiving individual speech sounds, a growing body of empirical work is focusing on connected speech. That issue is taken up here, outlining in particular the potential role of neuronal oscillations in analyzing speech. The brain basis of speech is, unsurprisingly, a central focus of much research in human hearing.

The other domain that is receiving a great deal of attention and generating a fascinating body of data concerns the cortical foundations of processing music, discussed by Zatorre and Zarate in Chapter 10. Three aspects of music processing receive special attention. First, the concept of pitch and its foundational role for melody and melodic processing is discussed from several vantage points. Second, some of the major anatomic issues are revisited, including the interesting hemispheric asymmetries associated with processing auditory signals as well as the controversial contribution of dorsal stream structures to auditory processing. Third, the fascinating issues surrounding cortical plasticity after training and cortical deficits after lesion/genetic anomaly are laid out. Music processing is a domain that, in an important sense, illustrates how different methodologies and different concepts (object, stream, action–perception loop, etc.) come together to highlight the neurobiological foundations of complex auditory processing.

Although this volume is about the human auditory cortex, it now goes without saying that there is a fundamental role for the other senses in auditory processing. The multisensory role of the human auditory cortex is the theme of Chapter 11. There, the neurophysiological foundations, to date largely based on results from animal studies, are outlined by van Wassenhove and Schroeder. Nevertheless, a growing literature in human cognitive neuroscience shows convincingly the direct and seemingly causal interactions in multisensory contexts. Several compelling psychophysical phenomena are explained, including ventriloquism and the famous audiovisual speech McGurk illusion. It is becoming increasingly clear that the notion of purely unisensory areas is highly problematic, if not entirely incorrect. The temporally (and anatomically) very early effects of multisensory influence and integration highlight the fact that it is very fruitful to incorporate the rich and highly modulatory multisensory inputs and their effects into any model.

In Chapter 12, an almost mythical area in human auditory neuroscience, the planum temporale, is discussed by Hickok and Saberi. When one talks about the neural basis of speech and language processing, there are two brain areas that invariably lie at the center of the discussion: Broca's region in the frontal lobe and planum temporale in the superior temporal lobe. These brain areas have been argued to be strongly lateralized in the human brain, and have been shown to correlate in important ways with properties of speech perception and production as well as language comprehension and production. Three aspects of planum temporale function are discussed in this chapter: its role in the analysis of spatial features of sounds and localization, its role for the analysis of objects, and its role in auditory motor mapping.

The volume concludes with Chapter 13, wherein Cariani and Micheyl summarize where we stand with respect to computational models that link the data from the multiple methodologies in human and animal models in an explicit way. Although our understanding of neural coding is incomplete and woefully inadequate, enough is known to begin to explore how the neural code forms the basis for auditory perception and cognition. This chapter serves to remind us, from many different angles, what some of the computational requirements and coding strictures are. To develop a theoretically well motivated, computationally explicit, and neurobiologically realistic model, linking hypotheses that are sensitive to the toolbox of computational neuroscience will be essential.

This volume constitutes a fair representation of what is currently known about human auditory cortex. If the topic is revisited in 10 years, where should the field be?

Given the fantastic rate of progress—in terms of computational sophistication, experimental subtlety, and methodological innovations with ever better resolution—here are some decadal desiderata. First, the anatomical characterization of human auditory cortex needs to be at least as granular as the specifications now available for, say, macaque visual cortex. Not just areal delineation, cytoarchitecture, and receptor distributions, of course, but circuit-level considerations are also necessary. Achieving this for the human brain in general, and for auditory cortex in particular, is a tremendous challenge.

Second, it would be invaluable to have an inventory of the elementary operations (or computations) that are executed by cell groups in a given anatomic circuit or region. In some sense, this amounts to identifying the "atoms of hearing." Ultimately, the goal will be to develop linking hypotheses between the anatomic circuitry and the computational primitives (or atoms). It is likely that such a mapping between anatomy and computational functions will be many to many; the structure of cortex is such that a subgroup of elements can execute different formal operations.

Third, it would constitute a great success to have a model of how the primitive operations—that are anatomically grounded—conspire to create the elementary perceptual attributes, as these are unlikely to be neurobiological primitives. For example, there are many ways to create the elementary perceptual experience of pitch, or of loudness, from which it follows that the underlying operations that form the basis for that experience are even more elementary. Some of the considerations outlined in the concluding chapter point toward general issues to think about in formulating such theories.

Fourth, none of this can be achieved without concomitant progress in advancing the resolution power of the available methodologies. It goes without saying that the loudness of the fMRI environment is extremely detrimental to auditory neuroscience. Presumably, were the scanner to produce visual noise, MRI technicians and researchers would have scrambled and found a way to reduce this by now (some 20 years after the conception of BOLD imaging); but because the scanner "only" produces acoustic noise, a significant reduction of the acoustic noise of fMRI (mainly due to hardware and acquisition techniques) remains the holy grail for auditory neuroscience.

Further, the simultaneous combination of techniques (e.g., EEG and fMRI) would provide an important step forward in elucidating the relationships between various aspects of the neural signature(s) in auditory cortex.

Finally, successful explanations of how the human auditory cortex provides the basis for the representation and processing of ecologically natural signals such as speech, music, or natural sounds need to be directly grounded in the anatomic and computational infrastructure (bottom-up constraints) while permitting the seamless integration with high-level representations, and the entire predictive machinery that is the memory system (top-down constraints). In summary, the systematic linking of a computational inventory with the right level of anatomic circuitry constitutes a goal that is ambitious but would yield great scientific payoff and have compelling clinical implications.

References

King, A. J., & Nelken, I. (2009). Unraveling the principles of auditory cortical processing: Can we learn from the visual system? *Nature Neuroscience*, 12(6), 698–701.

Manley, G. A., Fay, R. R., & Popper, A. N., Eds. (2008). *Active processes and otoacoustic emissions in hearing*. New York: Springer.

Meddis R., Lopez-Pevada, E., Fay, R. R., & Popper, A. N., Eds. (2010). *Computational models of the auditory system*. New York: Springer.

Oertel, D., Fay, R. R., & Popper, A. N., Eds. (2002). *Integrative functions in the mammalian auditory pathway*. New York: Springer.

Schnupp, J., Nelken, I., & King, A. (Eds.). (2011). *Auditory neuroscience: Making sense of sound*. Cambridge, MA: MIT Press.

Winer, J. A., & Schreiner, C. E., Eds. (2010). *The auditory cortex*. New York: Springer.

Part I
The Methods

Chapter 2
Architecture, Connectivity, and Transmitter Receptors of Human Auditory Cortex

Stephanie Clarke and Patricia Morosan

Abbreviations

A1	primary auditory area
AChE	acetylcholine esterase
CB	calbindin
CO	cytochrome oxidase
CR	calretinin
HG	Heschl's gyrus
PAC	primary auditory cortex
PP	planum polare
PT	planum temporale
PV	parvalbumin
STG	superior temporal gyrus
STS	superior temporal sulcus

2.1 Introduction

Human auditory cortex, located on the supratemporal plane, comprises in the vicinity of primary auditory cortex (PAC) several nonprimary auditory areas. Architectonic studies that benefited from methodological advances, such as observer-independent

S. Clarke (✉)
Service de Neurospychologie et de Neuroréhabilitation, CHUV,
1011 Lausanne, Switzerland
e-mail: Stephanie.Clarke@chuv.ch

P. Morosan
Institute of Neurosciences and Medicine (INM-1), Research Centre Jülich,
52425 Jülich, Germany
e-mail: p.morosan@fz-juelich.de

D. Poeppel et al. (eds.), *The Human Auditory Cortex*, Springer Handbook
of Auditory Research 43, DOI 10.1007/978-1-4614-2314-0_2,

analysis and functionally related stains, have identified specific areas whose involvement in speech analysis, sound recognition, and auditory spatial processing has been established in activation studies. Postmortem and in vivo tracing studies have revealed a complex pattern of intra- and interareal connections that partially resemble those described in nonhuman primates but that also display specifically human attributes. Current evidence reveals a model of parallel and hierarchical organization of the early-stage auditory areas with an early separation of specific processing streams.

2.2 Historic Concepts and Maps of Human Auditory Cortex

At the beginning of 20th century, Paul Flechsig identified the superior temporal gyrus (STG) as the cortical site of the human auditory system. By using a myelogenetic approach, Flechsig (1908) succeeded in tracking the auditory pathway from the thalamus to the upper bank of the STG. He also observed that a distinct region on the first transverse temporal gyrus, or Heschl's gyrus (HG), receives denser thalamic inputs from the medial geniculate body than the surrounding cortex. This region, initially called the "auditory sphere," is the PAC.

Flechsig's reports on local differences in anatomical connectivity shifted the focus of auditory research away from brain macroanatomy to the finer, microscopic details of cortical organization. The necessary methodological framework for a well-founded histological examination of brain tissue was then provided by Nissl (1894) and Weigert (1882), who introduced useful stains for demonstrating cell bodies and myelinated fiber tracts, respectively. The stained cells and fiber tracts appeared black against a very light background, and this high contrast encouraged many researchers to study the cellular (cyto-) and fiber (myelo-) architecture of human cerebral cortex.

Campbell (1905) was among the first to study the cyto- and myeloarchitecture of human auditory cortex. He identified an "audito-sensory" area on the upper bank of the STG that possessed architectonic features entirely different from those of any other part of the temporal lobe. According to Campbell (1905), the "audito-sensory" area was coextensive with Flechsig's "auditory sphere" and thus represented the architectonic correlate of the human PAC. Campbell also identified a second, nonprimary auditory or "audito-psychic" area, which mainly covered dorsocaudal and lateral portions of the superior temporal gyrus (Fig. 2.1A).

The most influential architectonic parcellation of human auditory cortex, however, was published few years later by Brodmann (1909). Brodmann, a co-worker of Cecile and Oskar Vogt at the Kaiser Wilhelm Institute in Berlin, confirmed the existence of an architectonically distinct PAC (area 41 according to Brodmann), but refined the concept of nonprimary auditory cortex by segregating it into two major areas, areas 42 and 22 (Fig. 2.1B). In addition, Brodmann identified a new area, area 52, at the medio-anterior border of area 41. Brodmann's research was based on the assumption that each architectonically distinct cortical area also differs in functionality. Although microstructure–function relationships in the human brain could not be rigorously tested at that time, old (Vogt & Vogt, 1919) and more recent (Luppino et al., 1991; Matelli et al., 1991) studies in nonhuman primates

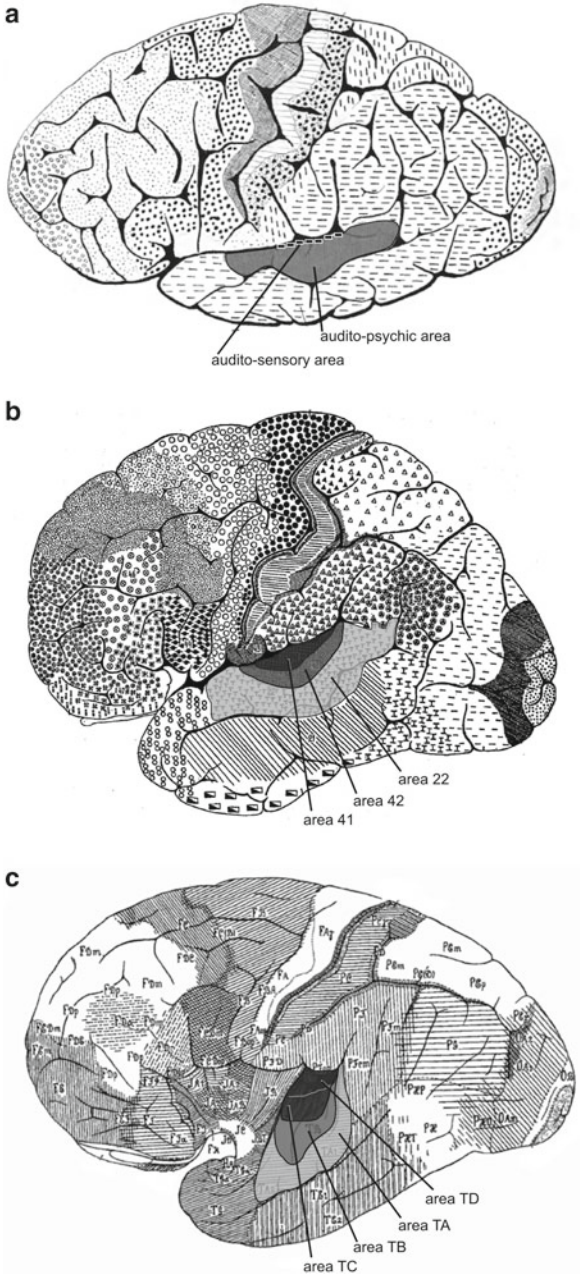


Fig. 2.1 Historic architectonic maps of human auditory cortex. Lateral view. (**a**) Myelo- and cytoarchitectonic map of Campbell (1905). (**b**) Cytoarchitectonic map of Brodmann (1909). (**c**) Cytoarchitectonic mapof von Economo and Koskinas (1925)

have demonstrated by means of combined electrophysiological–neuroanatomical studies that Brodmann's basic idea was true. Brodmann, however, did not argue for an extreme localization concept, that is, he did not try to relate complex function to one distinct architectonic area.

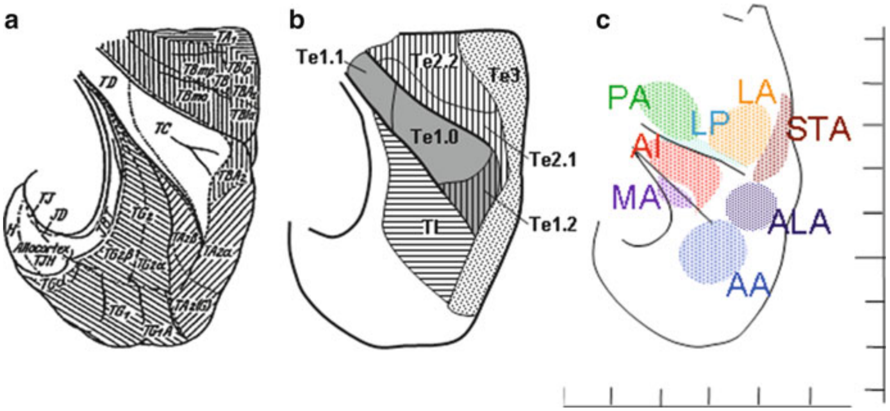


Fig. 2.2 Topography or primary and nonprimary auditory areas. (**a**) Classic cytoarchitectonic map of von Economo and Horn (1930). (**b**) Combined cyto- and receptor architectonic map (adapted from Morosan et al., 2001, 2005a). Areal borders were confirmed by using an algorithm-based, observer-independent method for detecting changes in cortical architecture. 3D stereotaxic probabilistic maps of the cortical areas and useful tools for anatomical localization of functional imaging data are available at http://www.fz-juelich.de/inm/index.php?index=397. (**c**) The primary (AI) and seven nonprimary auditory areas as identified histologically by Rivier and Clarke (1997) and Wallace et al. (2002), positioned within the Talairach coordinate system. All representations are upper views of the supratemporal plane with the temporal pole pointing down

A detailed and comprehensive description of the cytoarchitecture of human auditory cortex was published by von Economo and Koskinas (1925). The authors described meticulously, and partly quantitatively, the cytoarchitectonic properties of primary and nonprimary auditory areas, including information on topography, laminar dimensions, cell types, sizes, and densities. PAC (area TC according to von Economo and Koskinas [1925]) occupies central and anterior portions of HG, whereas the posterior portion of the gyrus contains area TD, a presumably additional primary auditory area (Fig. 2.1C). Areas TC and TD are bordered caudally by the nonprimary auditory area TB, which, in turn is bordered by area TA. The combined areas TC and TD correspond to Brodmann's area 41, whereas areas TB and TA resemble Brodmann's areas 42 and 22, respectively. More detailed topographic comparisons, however, reveal discrepancies between the two maps. In contrast to Brodmann's area 22, for instance, the nonprimary auditory area TA does not extend onto the middle temporal gyrus.

A few years later, von Economo and Horn (1930) published a much more complex cytoarchitectonic map of human auditory cortex, thus proposing that the structure of human auditory cortex is much more heterogeneous than initially believed (Fig. 2.2A). More significant, however, the analysis of a large number of brains (14 hemispheres) revealed striking intersubject and interhemispheric variations in the architecture and topography of auditory areas. These findings were confirmed by Sarkissov and colleagues (1955), who again used the terminology of Brodmann (1909).

One of the most important myeloarchitectonic maps of human auditory cortex was published by Hopf (1954). Hopf's map shows auditory cortex segregated into five major auditory areas: the highly myelinated PAC (or area ttr.1) and four nonprimary auditory areas (areas ttr.2, tpart, tsep, and tpari). Further, slight differences in myeloarchitecture enabled the segregation of each of those areas into a varying number of subareas. Area ttr1, for instance, was subdivided into four subareas along the anterior–posterior and themedial–lateral trajectories of Heschl's gyrus.

In the 1950s, the classic, purely subjective architectonic mapping strategies of human cerebral cortex begun to be subjected to careful and critical scrutiny (Bailey & von Bonin, 1951). Although the segregation of human auditory cortex into a primary and a nonprimary auditory region was still accepted, it has been argued that all other cortical subdivisions were not based on anatomical criteria that were objectively demonstrable. Indeed, the fundamental problem with classic architectonics was that many different criteria were used for parcellation, often introducing a major element of subjectivity in determining the areal borders, and thus the configuration of brain maps produced by different cartographers. It is, for example, generally accepted that the human PAC is confined to HG, but the exact position of the areal borders as well as the number and topographies of putative subdivisions remain a matter of debate (Campbell, 1905; Brodmann, 1909; von Economo & Koskinas, 1925; Beck, 1930; von Economo & Horn, 1930; Hopf, 1954, 1968; Sarkissov et al., 1955; Braak, 1978; Galaburda & Sanides, 1980; Ong & Garey, 1990; Rademacher et al., 1993; Rivier & Clarke, 1997; Clarke & Rivier, 1998; Hackett et al., 2001; Morosan et al., 2001; Wallace et al., 2002; Sweet et al., 2005; Fullerton & Pandya, 2007).

Several decades ago, however, Hopf (1968) introduced new quantitative techniques to describe the architecture of cortical areas and paved the way for modern, more objective and less observer-dependent architectonic mapping strategies of human auditory cortex (Schleicher et al., 1999, 2005). In addition, new mapping techniques have been developed that reflect functionally highly relevant information based on, for example, immunohistochemistry of transmitters and cytoskeletal elements and receptor autoradiography (Zilles et al., 2002b).

2.3 Primary Auditory Area

2.3.1 Relationship Between Heschl's Gyrus and Primary Auditory Cortex

HG as the cortical site of human PAC is an important, functionally relevant macroanatomical landmark of auditory cortex. PAC, however, is not coextensive with HG (von Economo & Horn, 1930; Rademacher et al., 1993, 2001; Morosan et al., 2001). Portions of PAC may surpass the framing sulci of HG and reach anteriorly the planum polare (PP) or posteriorly the planum temporale (PT). Equally possible, nonprimary auditory areas can partly extend on HG. In addition, the incidental occurrence of the intermediate sulcus complicates the cortical surface pattern of HG and its relationship to the architectonically defined PAC. This sulcus may mark the posterior border of PAC, but here again the overlap is far from perfect. Another critical region is the lateral border. The medio-to-lateral extent of PAC varies considerably between subjects, and the lateral flattening of HG or other canonical boundaries (Rademacher et al., 1993) are rather vague anatomic guides to the lateral end point of architectonically defined PAC.

Given that HG is a clearly visible anatomic structure at the spatial resolution of modern in vivo magnetic resonance (MR) imaging, it regularly serves as a structural marker for the localization of activation clusters obtained by functional neuroimaging. The discrepancies between HG and the architectonic borders of PAC, however, reveal that additional architectonic information is clearly needed for the definition of state-of-the-art structure–function relationships.

Moreover, it has been shown that the absolute size of PAC is grossly overestimated by any approach that interprets HG as the structural equivalent of PAC (Rademacher et al., 2001). This needs to be kept in mind when inferences about the size of PAC are made on the basis of in vivo HG volumetry (Penhune et al., 1996) or when gyral variations are taken as MR visible indicators of individual variations in physiology and behavior (Leonard et al., 1993; Rojas et al., 1997; Schneider et al., 2002; Warrier et al., 2009).

2.3.2 Architectonic Features of Primary Auditory Cortex

Human PAC has been repeatedly mapped on the basis of cortical architecture since the beginning of the last century (Campbell, 1905; Brodmann, 1909; von Economo & Koskinas, 1925; Beck, 1930; von Economo & Horn, 1930; Sarkissov et al., 1955; Hopf, 1954, 1968; Braak, 1978; Galaburda & Sanides, 1980; Ong & Garey, 1990; Rademacher et al., 1993; Rivier & Clarke, 1997; Clarke & Rivier, 1998; Hackett et al., 2001; Morosan et al., 2001; Wallace et al., 2002; Sweet et al., 2005; Fullerton & Pandya, 2007). The *koniocortical* appearance, that is, the predominance of small granular cells in all cortical layers, easily segregates it from the neighboring nonprimary auditory areas in cytoarchitectonic specimens (Campbell, 1905; von Economo & Koskinas, 1925; Galaburda & Sanides, 1980; Rademacher et al., 1993; Hackett et al., 2001; Morosan et al., 2001; Sweet et al., 2005; Fullerton & Pandya, 2007). The staining of layers II–IV appears dense and almost uniform and a slightly lighter stripe is found in lower layer V (Fig. 2.3A). The inner granular layer (layer IV) is generally well developed, presumably reflecting the dense thalamic inputs from the medial geniculate body targeting this layer. Layer III is populated by small to medium-sized pyramidal cells; larger neurons are rare. In strictly orthogonally cut brain sections, the small pyramidal cells of layer III are arranged in short radial columns, which partially extend into the neighboring cortical layers. This feature is usually referred to as the “rain shower formation” because it is reminiscent of fine, droplike laces (von Economo & Koskinas, 1925).

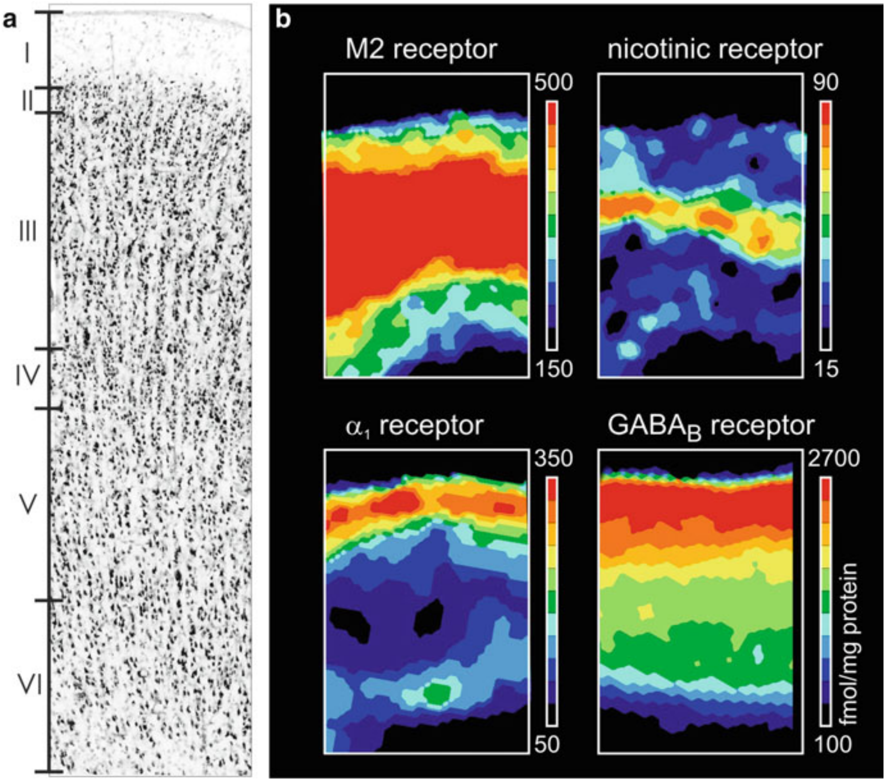


Fig. 2.3 Architecture of human primary auditory cortex (area Te1.0). (**a**) Cytoarchitecture. (**b**) Receptor architecture

In brain sections stained for myelin, primary auditory cortex attracts attention by its strong myelination in which radial fiber bundles can be followed from layer III to the white matter boundary (Campbell, 1905; Beck, 1930; Hopf, 1954; Hackett et al., 2001). The density of myelination is highest on the crown of HG (Hopf, 1954), and decreasing staining intensities have been observed from caudal to rostral portions of the gyrus (Hackett et al., 2001). The myeloarchitecture of PAC has been described as *astriate* (i.e., no horizontal stripes are visible in layers 4 or 5b due to almost uniformly dense fibrillarity from layer 4 through 6) (Hackett et al., 2001) to (*prope-*) *unistriate* (i.e., only layer 4 is visible due to relatively weaker myelination in layer 5a and uniformly dense staining of layers 5b and 6) (Hopf, 1954; Hackett et al., 2001). In this latter case, PAC is *internodensior* (i.e., layer 4 is less densely stained than layer 5b).

Chemoarchitectonically, PAC (area A1) is characterized by very high levels of acetylcholine esterase (AChE) and cytochrome oxidase (CO) activity (Hutsler & Gazzaniga, 1996; Rivier & Clarke, 1997; Clarke & Rivier, 1998; Hackett et al., 2001; Wallace et al., 2002; Sweet et al., 2005). In addition, it has very high densities of calcium binding proteins—parvalbumin (PV) and calretinin (CR) (Nakahara et al., 2000; Wallace et al., 2002; Chiry et al., 2003). AChE, CO, and PV have highest or very high densities in layer IV. AChE expression increases from layer IIIa to IV and decreases again in layers V and VI. AChE-positive pyramidal cells are rare throughout the entire cortical ribbon. In CO stains, layers I–III and V–VI are relatively light. PV labeling is less dark in layer III than in layer IV, and layers II, V, and VI are almost unlabeled. Most PV-positive elements, including neurons, were found in layers III and IV and the upper layer V. The CR neuropil labeling is light in layers II–III, absent in layer IV, and a dark stripe is again present in layer V. CR-positive neurons are found mostly in supragranular layers. Calbindin (CB), another calcium binding protein, forms a complementary laminar distribution pattern to that of PV. CB labeling is most intensive in layers II and IV, decreases in layer IV, and increases again in layer V.

In addition, human PAC has been mapped on the basis of multiple transmitter receptors (Zilles et al., 2002a; Morosan et al., 2005a) by using quantitative in vitro receptor autoradiography (Zilles et al., 2002a, b). Given that receptor molecules play a key role in cortical neurotransmission, analyzing their regional and laminar distribution patterns provides a direct link between functional and architectonic aspects of auditory cortical organization. The most conspicuous receptor architectonic features of human PAC are the extraordinarily high densities of cholinergic muscarinic M2 and nicotinic receptors, which reach maximum values in the midcortical layers (Fig. 2.3B). The M2 and nicotinic densities abruptly drop at the border to the nonprimary auditory areas. Like AChE, M2 and nicotinic receptors are tightly linked to cholinergic neurotransmission, and the presence of these three molecules at high levels in the thalamorecipient layers III/IV of PAC suggest a strong cholinergic modulation of human primary auditory signals. Together with the M2 and the nicotinic receptor, the γ-aminobutyric acid (GABA)ergic GABA$_A$, the noradrenergic α$_2$, and the serotonergic 5-HT$_2$ receptor subtypes also have higher densities in PAC than in the surrounding nonprimary auditory cortex and are denser in the midcortical than in the infragranular layers. Other transmitter receptors reach peak densities in the supragranular layers I–III. This set of receptors include the glutamatergic α-amino-3-hydroxyl-5-methyl-4-isoxazole-propionate (AMPA) and *N*-methyl-D-aspartate (NMDA) receptors, the GABAergic GABA$_B$ receptor, the cholinergic M1 and M3 receptors, the noradrenergic α$_1$ receptor, the serotonergic 5-HT$_{1A}$ receptor, and the dopaminergic D1 receptor (see Fig. 2.3B for the α$_1$ and GABA$_B$ receptors). The kainate receptor shows higher concentrations in layers V/VI and slightly higher values in layers I/II than in layers III/IV.

Recently, human PAC (called as area Te1, Fig. 2.2) and its related nonprimary auditory areas (called as areas Te2 and Te3) were mapped for the first time by using a novel, observer independent-method for localization of areal borders (Morosan et al., 2001, 2005a, b). This approach was based on the detection and localization of statistically significant changes in cortical (cyto-) architecture that occur at the border between two cortical areas (Schleicher et al., 1999, 2005). In brief, the cortical ribbon from layer II to the white matter boundary was systematically covered by a

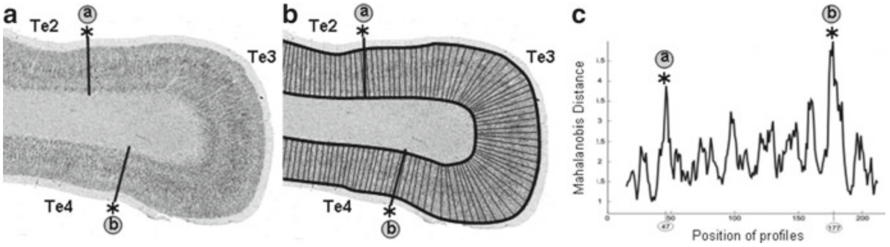


Fig. 2.4 Observer-independent localization of areal borders in cytoarchitectonic specimen. (**a**) Digitized region of interest showing the lateral convexity of the superior temporal lobe and the cytoarchitectonic borders between the nonprimary auditory areas Te2, Te3, and Te4. (**b**) The cortical ribbon covered by equidistant traverses. (**c**) Mahalanobis distance function. Significant maxima are indicated by (a) and (b). These maxima correspond to the borders between Te2/Te3 (a) and Te3/Te4 (b) as depicted in **a** and **b**

set of equidistant traverses (Fig. 2.4A). Each traverse defined the course of a cortical profile that captures the laminar distribution of cell volume densities. The cortical ribbon was then analyzed by using a sliding window procedure. The sliding window consisted of two abutting cortical sectors, each of which comprising a set on *n* cortical profiles. Moving these sliding windows across the cortical ribbon in one profile increments, Mahalanobis distances (MD) were calculated at each position. The values were added to an MD function, which described the dependency of the MD on the position of the sliding window (Fig. 2.4B). In this function, local maxima are found at positions where the cortical architecture changed abruptly, that is, where the sliding windows were centered over an areal border. Local maxima were subsequently tested for significance using a Hotelling's T^2-test corrected from multiple comparisons and considered as an indicator of an areal border only if they were found at comparable positions in at least three directly neighboring brain sections.

The full extent of area Te1 defined in this quantitative way is most closely comparable to that of the combined areas TD/TC of von Economo and Horn (1930) (Figs. 2.2A, B) and of areas 41 and 41/42 of Sarkissov et al. (1955). It also corresponds to area ttr.1 of Hopf (1954), the combined areas Kam/Kalt of Galaburda and Sanides (1980), and comprises area AI as defined by Rivier and Clarke (1997) (Table 2.1). In contrast to Brodmann's area 41, it never extends on the lateral bulge of the STG.

Human PAC is not an architectonically homogeneous cortical region (von Economo & Horn, 1930; Hopf, 1954; Galaburda & Sanides, 1980; Morosan et al., 2001, 2005a; Fullerton & Pandya, 2007). Central portions of PAC display most clearly koniocortical features and have the best developed layer IV. Medially, the cytoarchitecture of PAC is characterized by a "fogging" of the cortical layers, the increased number of medium-sized pyramidal cells in layer IIIc, and the less clear "rain shower formation." Laterally, the flattening of HG has cytoarchitectonic features of a "transitional field" between primary and nonprimary auditory cortex. By using the novel observer-independent method for the localization of areal borders (see earlier), area Te1 has been recently segregated into three cytoarchitectonically distinct subareas: Te1.0, Te1.1, and Te1.2 (Morosan et al., 2001) (Fig. 2.2B). The architectonic heterogeneity of PAC presumably enables an even more detailed parcellation of PAC, but to the present day no quantitative methods were applied to confirm the areal borders (Beck, 1930; von Economo & Horn, 1930; Hopf, 1954).

2.3.3 Intra-areal Compartments Within the Primary Auditory Area

The highly granular part of HG, which corresponds architectonically to area TC (von Economo & Koskinas, 1925)/Te1.0 (Morosan et al., 2001) (Fig. 2.2) and functionally to the primary auditory area, has been shown to comprise two anatomically defined intra-areal compartments (Clarke & Rivier, 1998). First, a 2.0–2.5 mm wide cortical band characterized by high levels of CO and AChE activity runs perpendicularly to the long axis of HG roughly at mid-distance between the most lateral and most medial limits of the primary auditory area. Second, a pattern of dark and light CO stripes is present in layers III and IV; they are approximately 500 μm wide and are oriented roughly in parallel to the long axis of HG.

Several studies revealed tonotopic maps on Heschl's gyrus, albeit with different orientation of the gradient. A map with low frequencies antero-laterally and high frequencies postero-medially was proposed on the basis of fMRI (Formisano et al., 2003; Talavage et al., 2004) and electrophysiological studies (Liégeois-Chauvel et al., 1991, 1994). An antero-posteriorly running gradient was visualized in two more recent studies, as well as two mirror-symmetric tonotopic maps within PAC in 7T fMRI (Humphries et al. 2010; 7T: Da Costa et al. 2011). Both models of tonotopic organization suggest that the anatomically defined wide CO band lies within the representation of frequencies that are relevant to speech processing. The functional significance of the thinner CO stripes that run parallel to the long axis of HG remains to be established.

2.4 Nonprimary Auditory Areas on the Supratemporal Plane

2.4.1 Cyto- and Myeloarchitecture

Nonprimary auditory cortex (nonPAC) lying directly behind PAC on the supratemporal plane has received by far the most attention in literature, not least because of its critical role in language-related processes. This cortical region occupies the largest portion (>90%) of the PT and corresponds to area 42 of Brodmann (1909), area TB of von Economo and Koskinas (1925) and von Economo and Horn (1930), and area

Table 2.1 Nomenclature of auditory areas

Mapping study	Heschl's gyrus	Temporal plane	Lateral STG	Polar plane
Brodmann (1909)	41	42	22	52
von Economo and Koskinas (1925)	TC/TD	TB	TA	IBT
Hopf (1954)	ttr1[a]	ttr2[a]	tsep.l/tpartr	tsep.m/tpari
Galaburda and Sanides (1980)	Kam/Kalt	PaAi, PaAe, PaAc, Tpt	PaAe, Tpt	ProA
Rivier and Clarke (1997)	AI	LA/PA	STA	MA
Morosan et al. (2005)	Te1 (Te1.1/ Te1.0/Te1.2)	Te2 (Te2.1/Te2.2)	Te3	TI

STG, superior temporal gyrus.
[a]and further subdivisions.

Te2 of Morosan et al. (2005b) (Figs. 2.1 and 2.2; Table 2.1). Heschl's sulcus anteriorly and the end of the horizontal ramus of the Sylvian fissure (SF) posteriorly may serve as its macroanatomical boundaries, but the overlap is far from perfect. The medial and the lateral borders are not marked by any gross anatomical landmarks. However, the architectonically defined nonprimary auditory region of the temporal plane does not extend onto insular cortex medially, and it does not occupy the convexity of the superior temporal gyrus laterally.

Architectonically, the nonPAC of the temporal plane takes an intermediate position between the PAC and the nonPAC region located on the lateral bulge of the superior temporal gyrus. The granular appearance of PAC is abandoned and a smattering of medium-sized pyramidal neurons occurs in layer IIIc. The cell density of layer V increases from PAC through the nonPAC region of the planum temporale to the nonPAC of the lateral STG (von Economo & Koskinas, 1925; Galaburda & Sanides, 1980; Morosan et al., 2005a), whereas myelin densities show an inverse gradient, with the highest values being found in PAC (Hopf, 1954). In strictly orthogonally orientated cytoarchitectonic brain sections, the pyramidal neurons of layer III (and neighboring layers) are arranged in vertical columns. These columns are referred to by von Economo and Koskinas (1925) as the "organ pipe formation." In myeloarchitectonic specimen, the laminar pattern is propeunistriate and commonly internodensior (i.e., layer 5b is more prominent than layer 4) (Hopf, 1954).

The architecture of the PT nonPAC is not uniform. Von Economo and Horn (1930) segregated this cortical region into anterior (TBa) and posterior (TBp) areas (Fig. 2.2A). In the posterior areas, layer III narrows in favor of layer V, which has also more densely packed neurons than the anterior area. The segregation of the PT along its anterio–posterior axis has also been reported in myeloarchitectonic studies (e.g., Hopf, 1954) and recently confirmed by observer-independent cytoarchitectonic and receptor mapping of areas Te2.1 and Te2.2 (Morosan et al., 2005a) (see Fig. 2.2B for Te2.1 and Te2.2).

The cortical region located anteriorly to PAC took a back seat to the posterior nonPAC in auditory research. This nonPAC region occupies large portions of the PP (von Economo & Koskinas, 1925; von Economo & Horn, 1930) and has been cytoarchitectonically characterized as different areas by various authors (see Table 2.1). Architectonically, this cortical region is characterized by a relative thin cortical ribbon and prominent infragranular layers.

2.4.2 Putative Functional Specialization of Nonprimary Auditory Areas

The heterogeneity of nonPAC has been investigated with functionally related stains by two independent groups (Rivier & Clarke, 1997; Wallace et al., 2002). Serial sections through the supratemporal plane were stained for Nissl, myelin, cytochrome oxidase, AChE, NADPH-diaphorase or parvalbumin; by cross-comparing the staining patterns, seven nonprimary areas, covering 1.1–3.1 cm^2 of cortical surface, have been identified and are referred to as PA (posterior auditory area), LA (lateral auditory area), LP (lateroposterior auditory area), ALA (anterolateral auditory area), AA (anterior auditory area), MA (medial auditory area), and STA (superior temporal auditory area; Fig. 2.2C). Further support for multiple auditory areas comes from calcium-binding protein expression patterns (parvalbumin, calbindin, calretinin; Wallace et al., 2002; Chiry et al., 2003), which greflect partially functional subdivisions within the subcortical auditory pathway and speak in favor of parallel ascending paths (Tardif et al., 2003). More recently, several GABA$_A$ and the two GABA$_B$ receptor subunits were shown to have a differential distribution within the supratemporal plane (Sacco et al., 2009). The characteristics of the staining patterns, such as prominence of a midcortical band in CO staining or the proportion of fiber versus somatic AChE staining, suggest that six of the nonprimary areas (PA, LP, LA, ALA, AA, MA) correspond to the same hierarchical level, while one area (STA) is at a higher level.

The supratemporal plane is involved in specific aspects of auditory analysis, including the recognition and localization of sound objects and the perception of speech. A meta-analysis of activation studies suggested that areas LA and STA play an important role in speech analysis (Scott & Johnsrude, 2003). Several other studies demonstrated that discrete regions of the supratemporal plane participate in sound recognition and sound localization (Griffiths & Warren, 2002; Hart et al., 2004; Hall et al., 2005; Barrett & Hall, 2006; Viceic et al., 2006; Altmann et al., 2007). Among areas known to be selective for sound recognition, ALA but not AA was shown to be modulated by the position of sound objects they code (Fig. 2.5; Van der Zwaag et al. 2011). The lateral part of the PT, including areas LA, STA, and PA, was shown to be involved equally in sound recognition and localization, whereas the medial part was more selectively involved in sound localization. Because of its involvement in processing of complex sounds and its interactions with higher order areas, the PT has been proposed to constitute a computational hub for both spatial

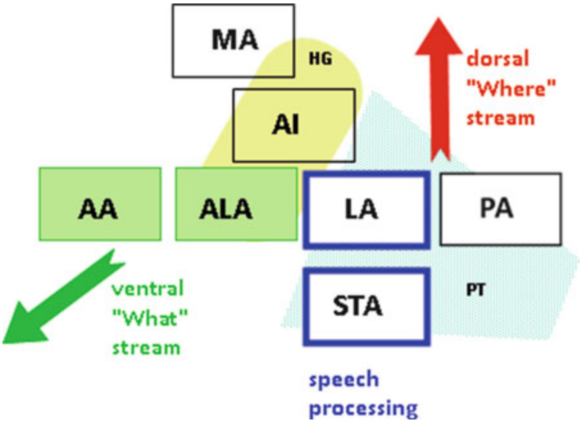


Fig. 2.5 Schematic representation of processing streams within the nonprimary auditory areas on the supratemporal plane. Comparison of histologically identified auditory areas (Rivier & Clarke, 1997; Wallace et al., 2002) with several activation studies (Griffiths & Warren, 2003; Scott & Johnsrude, 2003; Hart et al., 2004; Hall et al., 2005; Barrett & Hall, 2006; Viceic et al., 2006; Altmann et al., 2007) suggests that distinct processing streams originate in the supratemporal plane. In particular areas AA and ALA have been shown to be selectively activated by sound recognition as compared to sound localization (Viceic et al., 2006), participating thus in the putative auditory ventral or "What" stream. Areas LA, PA, and STA tended to be equally activated by recognition and localization (Viceic et al., 2006) and are most likely part of a computational hub for both spatial and nonspatial processing, which has been identified on the planum temporale (Griffiths & Warren, 2002); it is very likely that these areas contribute to the auditory dorsal or "Where" stream. Areas LA and STA were also shown to be involved in speech processing (Scott & Johnsrude, 2003); their position at the crossreads between the ventral and dorsal processing streams supports the dual-pathway for speech processing (Hickok & Poeppel, 2007). HG, Heschl's gyrus; PT, planum temporale

and nonspatial processing (Griffiths & Warren, 2002) (for discussion of some putative computational roles of the PT, see Hickok & Saberi, Chapter 12).

2.5 Temporal and Parietal Convexities

The temporal and parietal convexities contain higher auditory cortex. Although most researchers would relate activations of the free convexity of the superior temporal lobe to Brodmann's area 22, little is known about the architectonics and the topography of this cortical area. According to Brodmann's map (Brodmann, 1909), area 22 extends onto the posterior two thirds of the lateral bulge of the STG. Ventrally, it occupies large portions of the superior temporal sulcus region and the nearby middle temporal gyrus. The posterior border is marked by the vertical branch of the lateral fissure, while the anterior border may coincide with the crossing between a line continuing the course of the central sulcus and the lateral fissure. The position of the areal border within the depth of the superior temporal sulcus cannot be inferred from Brodmann's map.

Although the localization of an architectonically distinct area on the lateral bulge of the STG was confirmed by many investigators (von Economo & Koskinas, 1925; Hopf, 1954; Sarkissov et al., 1955; Galaburda & Sanides, 1980; Rivier & Clarke, 1997; Wallace et al., 2002), its exact position and extent are still a matter of debate. In particular, the precise location of the areal border within the superior temporal sulcus has remained loosely defined. Recently, the cortical region comprising BA22 was remapped on the basis of cyto- and receptor architectonics (Morosan et al., 2005b). The areal borders were detected by means of an observer-independent method, which was slightly modified to meet the needs of a combined cyto- and receptor architectonic mapping studies (Schleicher et al., 1999, 2005). Detailed comparisons of the newly defined cortical area (called area Te3; Fig. 2.2B) and previous architectonic maps revealed partly conflicting topographies. In contrast to BA22, Te3 did not extend on the middle temporal gyrus. The ventral border of Te3 and the multimodal area Te4 was consistently found on the upper bank of the superior temporal sulcus. Te3 thus did not occupy large portions of the STS region. Te3 is comparable with area TA of von Economo and Koskinas (1925) and von Economo and Horn (1930), area tsep.l and tpartr of Hopf (1954), and area PaAe of Galaburda and Sanides (1980) (see Table 2.1 this chapter). Unlike PaAe, however, area Te3 occupies posterior portions of the lateral bulge of the STG. Compared to the histochemically defined area STS of Rivier and Clarke (1997) and Wallace and colleagues (2002), Te3 extends to more rostral and more caudal levels, probably including brain regions that have yet not been mapped histochemically (Fig. 2.2B, C).

Cytoarchitectonically, Te3 is characterized by a prominent size and density of pyramidal neurons in deep layer III and a rather cell dense layer V. The granular layer II and IV are relative broad. Layer IV, however, is smaller than in primary auditory cortex. The neurons are arranged in vertical columns ("organ pipes" of von Economo & Koskinas, 1925). Unlike the neighboring areas, there is no sharp border between the cortex and the white matter. The cytoarchitectonically defined borders of Te3 closely matched changes in regional and laminar distribution patterns of various transmitter receptors. Te3 has lower densities of, for example, cholinergic muscarinic M2, glutamatergic NMDA, and GABAergic $GABA_A$ receptors than area Te2. Compared to the superior temporal sulcus area Te4, Te3 is characterized by, for example, lower densities of noradrenergic α_1, and of glutamatergic receptors.

The cytoarchitectonically defined temporoparietal area Tpt of Galaburda and Sanides (1980) is considered an additional field belonging to the auditory region in a broader sense. It occupies portions of the dorsolateral surface of the STG and extends toward the temporoparietal junction. Area Tpt corresponds in location to area TA_1 (von Economo & Koskinas, 1925) and Brodmann's area 22 in its posterior end. Layer IV of area Tpt is less well developed than in the neighboring nonprimary auditory areas and the cell density in layer V is relatively high. As a whole, area Tpt may represent a transitional field between the temporal and parietal cortices. The marked left–right differences in the amount of area Tpt were usually related to asymmetries of the language system (Galaburda et al., 1978).

2.6 Intersubject Variability and Probabilistic Mapping

Although reports on topographic variability of auditory areas date back to the 1930s (von Economo & Horn, 1930), it was only recently that the range of intersubject variations in area position, extent, and absolute size was quantified systematically (Galaburda & Sanides, 1980; Rademacher et al., 1993, 2001). The volume of cytoarchitectonically defined PAC, for instance, varies between 830 mm^3 and 2797 mm^3 in the left and between 787 mm^3 and 2787 mm^3 in the right hemisphere. In addition, the positions of areal borders show intersubject differences in the millimeter range (Rademacher et al., 2001). These findings indicate that state-of the-art architectonic maps should not display areal borders without giving at least an impression of their intersubject variability.

To this point, cytoarchitectonic probabilistic maps have been recently introduced (http://www.fz-juelich.de/ime/index.php?index=29). These maps are based on (1) algorithm-based definition of areal borders (Schleicher et al., 1999) in cell

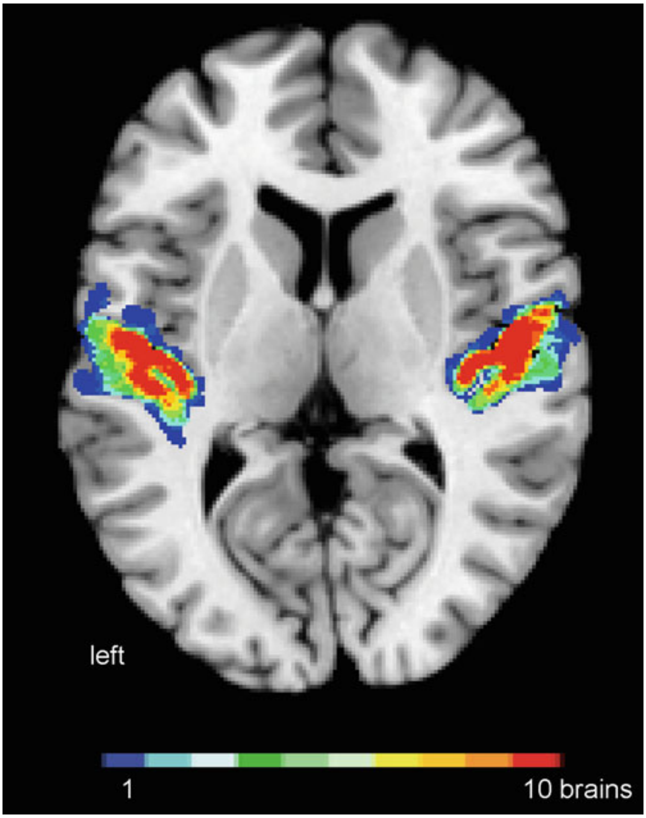


Fig. 2.6 Probability map of area Te1. red, highest probability; dark blue, lowest probability

body–stained sections of 10 postmortem brains, (2) 3D reconstruction of these sections using the MR data set of the same postmortem brain before its embedding in paraffin and sectioning, and (3) registration of these 3D data sets to a living standard references brain as common references space for cytoarchitectonic maps and functional imaging data (Zilles et al., 2002a; Amunts et al., 2006). Figure 2.6 presents a color-coded probabilistic map of PAC (as defined by the cytoarchitectonic area Te1) for the regions of identical probability in standardized space from 10% (dark blue) to 100% (red) (Rademacher et al., 2001; Morosan et al., 2005b). This map shows that the maximum probability of finding PAC is on the crown of HG and that the smaller the distance from the surface of HG to the fundus of its framing sulci, the greater the probability of finding nonprimary auditory areas. Given that exact and precise measurements of architectonic parcellation cannot—at least up to now—be made in living subjects, using the 3D probability map of PAC permits one to identify retrospectively the anatomical localization of significant cortical activations from functional studies with a distinct probability range. Such an approach will provide more accurate information on structure–function correlation in the human brain than the use of traditional architectonic maps that show schematically the parcellation scheme of a single brain.

2.7 Hemispheric Asymmetries

The macroanatomic geometry of the superior temporal lobe is asymmetric. The horizontal portion of SF is usually longer and more horizontally oriented in the left hemisphere than in the right (Eberstaller, 1890; Cunningham, 1892; Steinmetz et al., 1990; Ide et al., 1996; Jäncke & Steinmetz, 2004). SF asymmetries exist already at early stages of ontogeny (LeMay & Culebras, 1972) and appear to represent a conservative aspect of temporal lobe organization because it has been also observed in great apes (Zilles et al., 1996). In addition, the supratemporal plane (including HG and PT) is shifted rostrally in the right hemisphere relative to the left (Rademacher et al., 2001). The observation that the right superior temporal lobe is not a simple mirror image of its left counterpart advises caution when stereotaxic atlases depicting the macroanatomy of a single hemisphere (e.g., Talairach & Tournoux, 1988) are used as an anatomical reference for the localization of auditory clusters revealed by modern functional neuroimaging.

Initially reported by (Pfeifer, 1936), leftward asymmetries in the size of PT have been related to the typical left-hemisphere dominance for language functions (Geschwind & Levitsky, 1968). Currently, however, the asymmetry in PT size and its putative impact on left-lateralized language functions is highly disputed (Rubens et al., 1976; Loftus et al., 1993; Binder et al., 1996; Westbury et al., 1999). The PT asymmetry observed by earlier investigators appears to relate to the well known side differences in shape of the SF rather than to (functionally relevant) differences in PT size. Further, it has been argued that the reported PT size asymmetry (about 60%–80%) had been always much lower than the incidence of left-hemisphere language lateralization in the population (estimated to be >95%). Finally, structural asymmetries of auditory cortices appear not to relate to language lateralization as defined by the intracarotid sodium amytal testing (Dorsaint-Pierre et al., 2006).

Side differences have also been reported for HG (von Economo & Horn, 1930; Musiek & Reeves, 1990; Campain & Minckler, 1976; Kulynych et al., 1994; Penhune et al., 1996; Kennedy et al., 1998; Rademacher et al., 2001). In a large sample of postmortem brains (n = 27 brains), Rademacher and colleagues (2001) found a leftward asymmetry in almost half of the brains (48%), and an asymmetry favoring the right hemisphere in every fourth brain (25%). A larger HG volume in the left hemisphere than in the right (left: 1692 ± 659 mm^3; right: 1508 ± 342 mm^3) was confirmed by using in vivo morphometry (Penhune et al., 1996). The in vivo study also showed that the leftward HG asymmetry can rather be ascribed to differences in white than in gray matter. Despite initial reports, there is no consistent constellation with a single HG found predominantly in the left hemisphere and HG duplications found in the right hemisphere (von Economo & Horn, 1930; Campain & Minckler, 1976; Musiek & Reeves, 1990; Penhune et al., 1996; Leonard et al., 1998; Rademacher et al., 2001).

The macroanatomic asymmetries of HG and PT are influenced by genetic and/or epigenetic factors. Reduced PT asymmetries, for instance, were found in healthy left-handers when compared to right-handers (Steinmetz, 1996; Jäncke & Steinmetz, 2004) and in brains of patients suffering from schizophrenia (Chance et al., 2008), whereas musicians with perfect pitch (those who identify any musical tone without a reference tone) had a PT asymmetry twice as great as nonmusicians and musicians without perfect pitch (Schlaug et al., 1995). Similarly, musicians show larger HG gray matter than nonmusicians across hemispheres (Schneider et al., 2002, 2005).

Little is known about putative differences between left and right auditory cortices at the microstructural level. Asymmetries have been reported in, for example, the sizes of architectonically defined areas (Galaburda & Sanides, 1980), the spacing of cell columns (Seldon, 1981), the size of IIIc pyramidal neurons (Hutsler & Gazzaniga, 1996), and the depth of layer IV (Morosan et al., 2001).

2.8 Connectivity of Auditory Cortex

The connectivity of human auditory cortex is often presumed to be very similar to that of nonhuman primates. This is probably true for the overall pattern of connectivity, including that of the early stage auditory areas. However, human auditory cortex is closely associated with lateralized functions, such as language (e.g., Scott & Johnsrude, 2003) or spatial cognition (Spierer et al., 2009), and may thus have very specific connectivity patterns involving higher-order cognitive areas.

The auditory information is conveyed to the primary auditory area via the massive thalamic input, as shown by studies on evoked potentials (Liegeois-Chauvel et al., 1991) and evoked magnetic fields (Romani et al., 1982; Hari et al., 1984; Yamamoto et al., 1988). This input originates most likely in the parvocellular subdivision of the medial geniculate nucleus (Dejerine & Dejerine-Klumpke, 1901; Locke et al., 1962; Van Buren & Borke, 1972) and travels though the acoustic radiation (Pfeifer, 1920). A recent postmortem study using myelin staining has mapped stereometrically the trajectory of the acoustic radiation and shown considerable interindividual and interhemispheric variability as to its precise position and volume (Rademacher et al., 2002). Although the primary auditory area tends to be asymmetrical (Penhune et al., 1996), this was shown not to be the case for the acoustic radiation or the medial geniculate body (Rademacher et al., 2002).

The intrinsic connectivity of the human primary and early-stage nonprimary auditory areas has been investigated with the anterograde and retrograde tracer DiI (Tardif & Clarke, 2001; Fig. 2.7). The intrinsic connections originate mostly from layer II–III pyramids; at short distances they spread densely in all cortical layers, but at longer distances they are less present in layer IV. Within the primary auditory area and the medially adjacent part of HG (area TD), the intrinsic connections involve a relatively narrow part of cortex. They spread over larger parts of cortex in the nonprimary areas on the plana polare and temporal (areas TG, TA, and TB), where they also tend to have anisotropic distributions. The observed differences suggest that intrinsic connections play a different role in primary and nonprimary auditory areas. Within the primary area, they are likely to involve nearby units or modules, probably with similar coding properties, whereas in nonprimary areas they spread most likely over more distant units and may play an important role in the integration of different auditory features.

The intra- and interareal connectivity of the primary auditory area and another tonotopically organized area on the lateral part of HG (most likely ALA or area Te1.2) has been investigated in vivo with DTI (Upadhyay et al., 2007). Intra-areal connectivity between representations of different frequencies was found to be stronger than interareal connectivity between tonotopically corresponding parts. This connectivity pattern resembles that of macaque core areas AI and R (Morel et al., 1993).

Peroperative recordings of evoked potentials further support close connectivity between the human primary area and the early-stage nonprimary auditory areas. The differences between response latencies to auditory stimuli in these areas are relatively small; latencies were reported to be 30–50 ms in AI and 60–75 ms in the more lateral part of Heschl's gyrus (Liégeois-Chauvel et al., 1994; the latter area is likely to correspond to ALA/Te1.2). Further, electrical stimulations in the primary auditory area yielded neural responses in the surrounding nonprimary cortex with short latencies, which were 6–8 ms in the anterolateral part of Heschl's gyrus and lateral to it (Liégeois-Chauvel et al., 1991; the recorded sites corresponded most likely to ALA/Te1.2 and PA/Te2) or 2–3 ms in the upper part of the superior temporal gyrus (Howard et al., 2000; the recorded site corresponded most likely to STA/Te3).

These observations from human studies are compatible with the connectivity pattern described in nonhuman primates. Auditory cortex of nonhuman primates is subdivided into core, containing the primary areas AI and R; the belt, which flanks the core laterally and medially and that contains areas AL, ML, CL, CM, and RM; and the parabelt, which is lateral to the (lateral) belt and is subdivided into rostral and caudal parts. Tracing studies have demonstrated that the primary auditory area

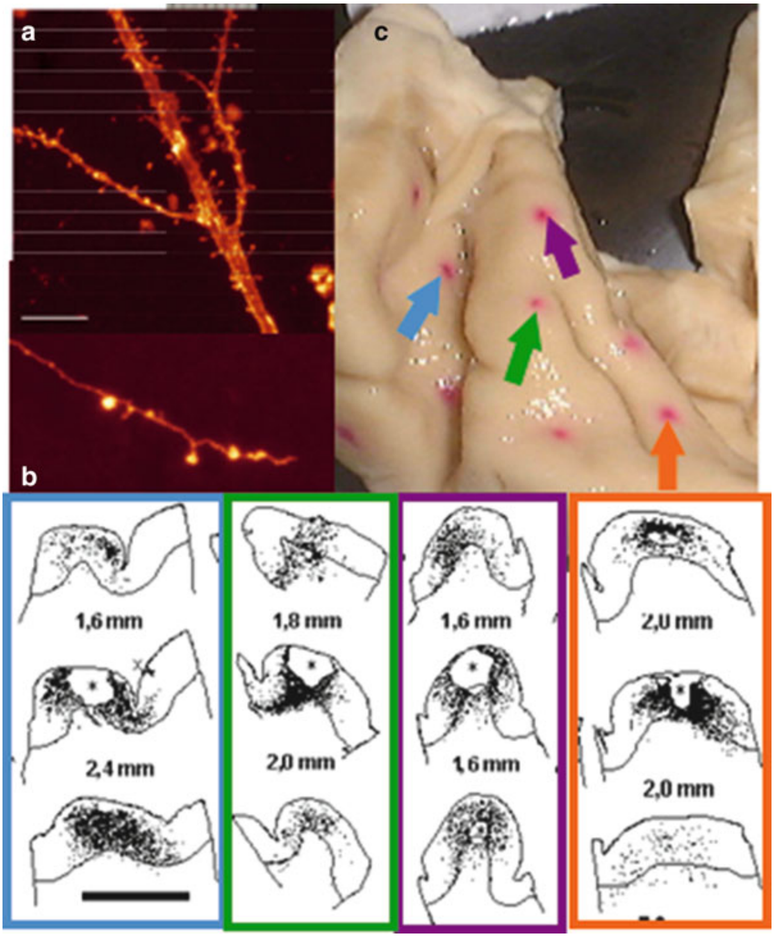


Fig. 2.7 Intrinsic connections of the human primary and nonprimary auditory areas as demonstrated with anterograde and retrograde labeling with DiI. (**a**) Apical dendritic arbor of a retrogradely labeled pyramidal neuron in layer III. (**b**) Labeled axon with terminal and "en passant" boutons. (**c**) Placement of injections on the supratemporal plane and charts showing the distribution of labeled neurons and axons in the primary auditory area (green) and areas TC (mauve), TB (blue), and TG (orange). Representative sections from serial charts are shown; for each reconstruction, medial sections are on top and lateral ones at the bottom of the column. For individual sections posterior is to the right and dorsal up. The mediolateral distance between shown sections is indicated in mm. Measure bars = 10 µm (**a, b**) and 5 mm (**c**). (Adapted from Tardif and Clarke, 2001.)

has strong connections with the other core area as well as with the belt, but not with parabelt areas; the belt areas have strong connections with each other, with the core and with the parabelt; and the parabelt has strong connections with the belt, but not the core (e.g., Morel et al., 1993; Hackett et al., 1998; de la Mothe et al., 2006). The parabelt and the adjacent superior temporal gyrus are connected with prefrontal cortex (e.g., Hackett et al., 1999a; Romanski et al., 1999) and receive in their caudal parts somatosensory input (Hackett et al., 2007; Smiley et al., 2007).

To summarize, current evidence from nonhuman primates suggests that there is: (1) a cascade of connections from core to belt to parabelt; (2) multisensory input to caudal parabelt; and (3) long-range connections between parabelt and prefrontal cortex. In humans, there is evidence for a similar type of connectivity within early-stage auditory areas; the connectivity of higher-order auditory and related cortex appears to follow the blueprint of nonhuman primates, however, with several notable differences.

The connectivity outside the human early-stage auditory areas shows more complex patterns. This is the case of the temporal part of Wernicke's area (BA 22) and its right homologue, where intrinsic connections have been investigated with the retrograde and anterograde tracers DiI and DiA (Galuske et al., 2000). Retrogradely labeled neurons were distributed in clusters; the spacing of the clusters, but not their diameter, was found to be larger in the left than the right hemisphere. Another example of the complex connectivity pattern in human cortex is Broca's area and its right homologue, which receives input from auditory association cortex (Parker et al., 2005; Hagmann et al., 2006; Barrick et al., 2007; Gharabaghi et al., 2009); its intrinsic connections have been investigated with the anterograde and retrograde tracers DiI and BDA (Tardif et al., 2007). They were shown to spread much more widely than in auditory cortex and to be layer specific. At short range they involve all cortical layers, but remain laminar specific at long range and clustered in supragranular layers.

The human long-range intrahemispheric connections have been investigated more recently in vivo using diffusion tensor imaging. Wernicke's and Broca's areas and their homologues in the right hemisphere were shown to be linked by two intrahemispheric auditory-language pathways; the dorsal pathway takes the arcuate fasciculus and the ventral the external capsula and the uncinate fasciculus (Parker et al., 2005). The connections between Broca's and Wernicke's areas are stronger than those between their homologues on the right side (Parker et al., 2005; Hagmann et al., 2006; Barrick et al., 2007; Gharabaghi et al., 2009), but the asymmetry within these pathways appears to be modulated by hand preference and sex (Hagmann et al., 2006). Another white matter pathway linking the posterior temporal lobe to the superior parietal lobule and putatively involved in audition has been shown to be stronger on the right side (Barrick et al., 2007); this rightward asymmetry may be linked to the predominant role the right hemisphere plays in auditory spatial processing (e.g., Spierer et al., 2009).

By analogy to nonhuman primates (e.g., Hackett et al., 1999b), it is generally assumed that human auditory areas have interhemispheric connections. However—and unlike for the visual callosal connections (Clarke & Miklossy, 1990; Clarke, 1994)—there is currently no direct evidence for homotopic callosal connections between the human auditory cortices, due to methodological difficulties. Two tracing studies using the Nauta method for anterogradely degenerating axons suggest, however, that the interhemispheric connectivity of human auditory cortex is likely to be complex. In a first study, it was shown that the right fusiform gyrus sends

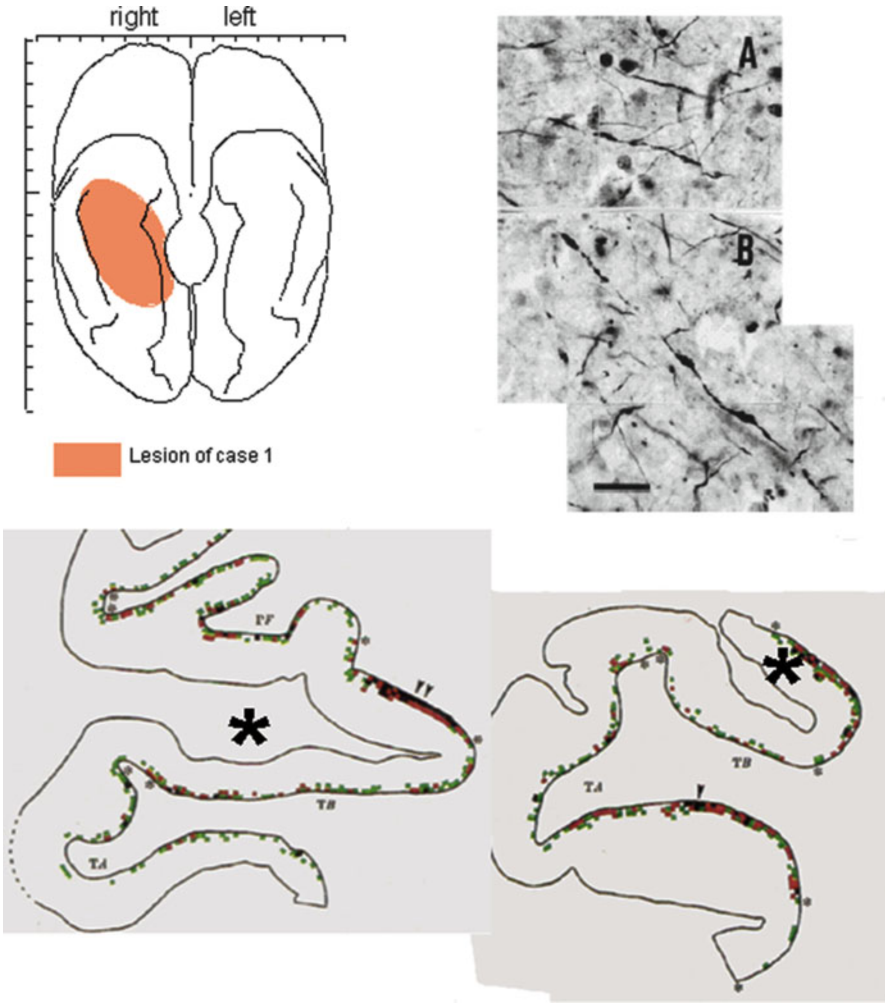


Fig. 2.8 Heterotopic interhemispheric connections to the supratemporal plane. The Nauta method for anterogradely degenerating axons has been used in a case with a unique focal lesion within the right fusiform and parahippocampal gyri (top left). Anterogradely degenerating axons (top right) were present in large parts of contralateral cortex, including the (left) planum temporale (adapted from Di Virgilio & Clarke, 1997). Asterisks mark the planum temporale; the color squares denote the density of degenerating axon segments within a 600-μm thick strip of cortex along the layer VI–white matter border

monosynaptic input to Broca's and Wernicke's areas, including the planum temporale and the posterior part of the superior temporal gyrus (Di Virgilio & Clarke, 1997; Fig. 2.8); no such input was present within the primary auditory area (Di Virgilio and Clarke, unpublished results). Thus, the nonprimary auditory areas on the planum temporale receive visual input, which is in keeping with the multisensory input to the caudal parabelt areas of nonhuman primates (Smiley et al., 2007). The widespread heterotopic interhemispheric connectivity appears, however, to be a human characteristic, which contrasts with the topographically rather precise corticothalamic (Clarke et al., 1999), corticostriatal (Wiesendanger et al., 2004), and collicular connectivity (Tardif & Clarke, 2002; Tardif et al., 2005). A second study, focusing on the anterior commissure, suggests a further human versus nonhuman difference in the organization in the interhemispheric connections. Using the Nauta method, it demonstrated a strong contribution from the anterior and inferior temporal lobe to the anterior commissure, similar to that described in nonhuman primates, but also a weaker contribution from other cortical areas, including the central fissure and dorsal prefrontal cortex (Di Virgilio et al., 1999). Thus, the anterior commissure may convey axons from large parts of human cortex, probably larger parts than in nonhuman primates.

2.9 Summary

Throughout the 20th century numerous architectonic studies revealed the structural complexity of human auditory cortex. Based on observer-independent methods and on functionally related stains, current evidence reveals that the PAC, located on Heschl's gyrus, is surrounded by at least seven other, nonprimary auditory areas. Comparison of histological and activation studies suggests that specific nonprimary areas are involved in speech analysis, sound recognition, and/or auditory spatial processing and may thus be at the origin of specialized processing pathways.

The connectivity of human auditory cortex is often presumed to be very similar to that of nonhuman primates. Evidence from human electrophysiological and tracing studies speaks in favor of a similar type of connectivity within early-stage auditory areas, whereas the connectivity of higher-order auditory and related cortex shows several notable differences.

Anatomical studies provide a key to the functional organization of human auditory cortex. In the absence of clear or easily demonstrable tonotopic organization, the identification of human auditory areas relies on histological investigations and on the cross-comparison with activation studies. As in nonhuman primates, current evidence suggests that parallel streams are devoted to the processing of recognition and spatial aspects, respectively (e.g., Maeder et al., 2001; Clarke et al., 2002). Specific human nonprimary areas are, however, also involved in speech processing and may be at the origin of the dual speech processing pathway (Hickok & Poeppel, 2007).

Acknowledgments This work has been supported by the Swiss National Science (Grantt 3200030-124897 to S. Clarke) and the Federal Ministry of Education and Research (BMBF Grant 01GW0623 to P. Morosan).

References

Altmann, C. F., Bledowski, C., Wibral, M., & Kaiser, J. (2007). Processing of location and pattern changes of natural sounds in the human auditory cortex. *NeuroImage*, 35, 1192–1200.

Amunts, K., Schleicher, A., & Zilles, K. (2006). Cytoarchitecture of the cerebral cortex – More than localization. *NeuroImage*, 37, 1061–1065.

Bailey, P., & von Bonin, G. (1951). *The isocortex of man*. Urbana: University of Illinois Press.

Barrett, D. J., & Hall, D. A. (2006). Response preferences for "what" and "where" in human non-primary auditory cortex. *NeuroImage*, 32, 968–977.

Barrick, T. R., Lawes, I. N., Mackay, C. E., & Clark, C. A. (2007). White matter pathway asymmetry underlies functional lateralization. *Cerebral Cortex*, 17, 591–598.

Beck, E. (1930). Die Myeloarchitektonik der dorsalen Schläfenlappenrinde beim Menschen. *Journal für Psychologie und Neurologie*, 41, 129–263.

Binder, J. R., Frost, J. A., Hammeke, T. A., Rao, S. M., & Cox, R. W. (1996). Function of the left planum temporale in auditory and linguistic processing. *Brain*, 119(4), 1239–1247.

Braak, H. (1978). The pigment architecture of the human temporal lobe. *Anatomy and Embryology*, 154(2), 213–240.

Brodmann, K. (1909). *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Leipzig: Barth.

Campain, R., & Minckler, J. (1976). A note on the gross configurations of the human auditory cortex. *Brain and Language*, 3, 318–323.

Campbell, A. W. (1905). *Histological studies on the localization of cerebral function*. Cambridge, UK: Cambridge University Press.

Chance, S. A., Casanova, M. F., Switala, A. E., & Crow, T. J. (2008). Auditory cortex asymmetry, altered minicolumn spacing and absence of ageing effects in schizophrenia. *Brain*, 131, 3178–3192.

Chiry, O., Tardif, E., Magistretti, P. J., & Clarke, S. (2003). Patterns of calcium-binding proteins support parallel and hierarchical organization of human auditory areas. *European Journal of Neuroscience*, 17, 397–410.

Clarke, S. (1994). Association and intrinsic connections of human extrastriate visual cortex. *Proceedings of the Royal Society London B: Biological Sciences*, 257, 87–92.

Clarke, S., & Miklossy, J. (1990). Occipital cortex in man: Organization of callosal connections, related myelo- and cytoarchitecture, and putative boundaries of functional visual areas. *Journal of Comparative Neurology*, 298, 188–214.

Clarke, S., & Rivier, F. (1998). Compartments within human primary auditory cortex: Evidence from cytochrome oxidase and acetylcholinesterase staining. *European Journal of Neuroscience*, 10, 741–745.

Clarke, S., Riahi-Arya, S., Tardif, E., Eskenasy, A. C., & Probst, A. (1999). Thalamic projections of the fusiform gyrus in man. *European Journal of Neuroscience*, 11, 1835–1838.

Clarke, S., Bellmann, T. A., Maeder, P., Adriani, M., Vernet, O., Regli, L., et al. (2002). What and where in human audition: Selective deficits following focal hemispheric lesions. *Experimental Brain Research*, 147, 8–15.

Cunningham, D. J. (1892). *Contribution to the surface anatomy of the cerebral hemispheres: Cunningham memoirs*. Dublin: Royal Irish Academy.

Da Costa, S., van der Zwaag, W., Marques, J. P., Frackowiak, R. S. J., Clarke, S., & Saenz, M. (2011). Human primary auditory cortex follows the shape of Heschl's gyrus. *Journal of Neuroscience*, 31, 14067–14075.

Dejerine, J., & Dejerine-Klumpke, A. (1901) Anatomie des Centres Nerveux.Tome II. Paris: J. Rueff.

de la Mothe, L. A., Blumell, S., Kajikawa, Y., & Hackett, T. A. (2006). Thalamic connections of the auditory cortex in marmoset monkeys: Core and medial belt regions. *Journal of Comparative Neurology*, 496, 72–96.

Di Virgilio G., & Clarke, S. (1997). Direct interhemispheric visual input to human speech areas. *Human Brain Mapping*, 5, 347–354.

Di Virgilio G., Clarke, S., Pizzolato, G., & Schaffner, T. (1999). Cortical regions contributing to the anterior commissure in man. *Experimental Brain Research*, 124, 1–7.

Dorsaint-Pierre, R., Penhune, V. B., Watkins, K. E., Neelin, P., Lerch, J. P., Bouffard, M. et al. (2006). Asymmetries of the planum temporale and Heschl's gyrus: Relationship to language lateralization. *Brain*, 129, 1164–1176.

Eberstaller, O. (1890). *Das Stirnhirn: Ein Beitrag zur Anatomie der Oberfläche des Gehirns*. Wien: Urban & Schwarzenberg.

Flechsig, P. (1908). Bemerkungen über die Hörsphäre des menschlichen Gehirns. *Neurologisches Zentralblatt*, 27, 2–7.

Formisano, E., Kim, D. S., Di, S. F., van de Moortele, P. F., Ugurbil, K., & Goebel, R. (2003). Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron*, 40, 859–869.

Fullerton, B. C., & Pandya, D. N. (2007). Architectonic analysis of the auditory-related areas of the superior temporal region in human brain. *Journal of Comparative Neurology*, 504, 470–498.

Galaburda, A., & Sanides, F. (1980). Cytoarchitectonic organization of the human auditory cortex. *Journal of Comparative Neurology*, 190, 597–610.

Galaburda, A. M., Sanides, F., & Geschwind, N. (1978). Human brain. Cytoarchitectonic left-right asymmetries in the temporal speech region. *Archives of Neurology*, 35, 812–817.

Galuske, R. A., Schlote, W., Bratzke, H., & Singer, W. (2000). Interhemispheric asymmetries of the modular structure in human temporal cortex. *Science*, 289, 1946–1949.

Geschwind, N., & Levitsky, W. (1968). Human brain: Left-right asymmetries in temporal speech region. *Science*, 161, 186–187.

Gharabaghi, A., Kunath, F., Erb, M., Saur, R., Heckl, S., Tatagiba, M., et al. (2009). Perisylvian white matter connectivity in the human right hemisphere. *BMC Neuroscience*, 10, 15.

Griffiths, T. D., & Warren, J. D. (2002). The planum temporale as a computational hub. *Trends in Neurosciences*, 25, 348–353.

Hackett, T. A., Stepniewska, I., & Kaas, J. H. (1998). Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *Journal of Comparative Neurology*, 394, 475–495.

Hackett, T. A., Stepniewska, I., & Kaas, J. H. (1999a). Prefrontal connections of the parabelt auditory cortex in macaque monkeys. *Brain Research*, 817, 45–58.

Hackett, T. A., Stepniewska, I., & Kaas, J. H. (1999b). Callosal connections of the parabelt auditory cortex in macaque monkeys. *European Journal of Neuroscience*, 11, 856–866.

Hackett, T. A., Preuss, T. M., & Kaas, J. H. (2001). Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *Journal of Comparative Neurology*, 441, 197–222.

Hackett, T. A., Smiley, J. F., Ulbert, I., Karmos, G., Lakatos, P., de la Mothe, L. A., et al. (2007). Sources of somatosensory input to the caudal belt areas of auditory cortex. *Perception*, 36, 1419–1430.

Hagmann, P., Cammoun, L., Martuzzi, R., Maeder, P., Clarke, S., Thiran, J. P., et al. (2006). Hand preference and sex shape the architecture of language networks. *Human Brain Mapping*, 27, 828–835.

Hall, D. A., Barrett, D. J., Akeroyd, M. A., & Summerfield, A. Q. (2005). Cortical representations of temporal structure in sound. *Journal of Neurophysiology*, 94, 3181–3191.

Hari, R., Hamalainen, M., Ilmoniemi, R., Kaukoranta, E., Reinikainen, K., Salminen, J., et al. (1984). Responses of the primary auditory cortex to pitch changes in a sequence of tone pips: Neuromagnetic recordings in man. *Neuroscience Letters*, 50, 127–132.

Hart, H. C., Palmer, A. R., & Hall, D. A. (2004). Different areas of human non-primary auditory cortex are activated by sounds with spatial and nonspatial properties. *Human Brain Mapping*, 21, 178–190.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8, 393–402.

Hopf, A. (1954). Die Myeloarchitektonik des Isocortex temporalis beim Menschen. *Journal für Hirnforschung*, 1, 208–279.

Hopf, A. (1968). Photometric studies on the myeloarchitecture of the human temporal lobe. *Journal für Hirnforschung*, 10, 285–297.

Howard, M. A., Volkov, I. O., Mirsky, R., Garell, P. C., Noh, M. D., Granner, M., et al. (2000). Auditory cortex on the human posterior superior temporal gyrus. *Journal of Comparative Neurology*, 416, 79–92.

Humphries, C., Liebenthal, E., & Binder, J. R. (2010). Tonotopic organization of human auditory cortex. *NeuroImage*, 50, 1202–1211.

Hutsler, J. J., & Gazzaniga, M. S. (1996). Acetylcholinesterase staining in human auditory and language cortices: Regional variation of structural features. *Cerebral Cortex*, 6, 260–270.

Ide, A., Rodriguez, E., Zaidel, E., & Aboitiz, F. (1996). Bifurcation patterns in the human sylvian fissure: Hemispheric and sex differences. *Cerebral Cortex*, 6, 717–725.

Jäncke, L., & Steinmetz, H. (2004). Anatomical brain asymmetries and their relevance for functional asymmetries. In K. Hughdahl & R. J. Davidson (Eds.), *The asymmetrical brain* (pp. 187–229). Cambridge, MA: MIT Press.

Kennedy, D. N., Lange, N., Makris, N., Bates, J., Meyer, J., & Caviness, V. S., Jr. (1998). Gyri of the human neocortex: An MRI-based analysis of volume and variance. *Cerebral Cortex*, 8, 372–384.

Kulynych, J. J., Vladar, K., Jones, D. W., & Weinberger, D. R. (1994). Gender differences in the normal lateralization of the supratemporal cortex: MRI surface-rendering morphometry of Heschl's gyrus and the planum temporale. *Cerebral Cortex*, 4, 107–118.

LeMay, M., & Culebras, A. (1972). Human brain—morphologic differences in the hemispheres demonstrable by carotid arteriography. *New England Journal of Medicine*, 287, 168–170.

Leonard, C. M., Voeller, K. K., Lombardino, L. J., Morris, M. K., Hynd, G. W., Alexander, A. W., et al. (1993). Anomalous cerebral structure in dyslexia revealed with magnetic resonance imaging. *Archives of Neurology*, 50, 461–469.

Leonard, C. M., Puranik, C., Kuldau, J. M., & Lombardino, L. J. (1998). Normal variation in the frequency and location of human auditory cortex landmarks. Heschl's gyrus: Where is it? *Cerebral Cortex*, 8, 397–406.

Liégeois-Chauvel, C., Musolino, A., & Chauvel, P. (1991). Localization of the primary auditory area in man. *Brain*, 114(1A), 139–151.

Liégeois-Chauvel, C., Musolino, A., Badier, J. M., Marquis, P., & Chauvel, P. (1994). Evoked potentials recorded from the auditory cortex in man: Evaluation and topography of the middle latency components. *Electroencephalography and Clinical Neurophysiology*, 92, 204–214.

Locke, S., Angevine, J. B., Jr., & Marin, O. S. (1962). Projection of the magnocellular medical geniculate nucleus in man. *Brain*, 85, 319–330.

Loftus, W. C., Tramo, M. J., Thomas, C. E., Green, R. L., Nordgren, R. A., & Gazzaniga, M. S. (1993). Three-dimensional quantitative analysis of hemispheric asymmetry in the human superior temporal region. *Cerebral Cortex*, 3, 348–355.

Luppino, G., Matelli, M., Camarda, R. M., Gallese, V., & Rizzolatti, G. (1991). Multiple representations of body movements in mesial area 6 and the adjacent cingulate cortex: An intracortical microstimulation study in the macaque monkey. *Journal of Comparative Neurology*, 311, 463–482.

Maeder, P. P., Meuli, R. A., Adriani, M., Bellmann, A., Fornari, E., Thiran, J. P., et al. (2001). Distinct pathways involved in sound recognition and localization: A human fMRI study. *NeuroImage*, *14*, 802–816.

Matelli, M., Luppino, G., & Rizzolatti, G. (1991). Architecture of superior and mesial area 6 and the adjacent cingulate cortex in the macaque monkey. *Journal of Comparative Neurology*, 311, 445–462.

Morel, A., Garraghty, P. E., & Kaas, J. H. (1993). Tonotopic organization, architectonic fields, and connections of auditory cortex in macaque monkeys. *Journal of Comparative Neurology*, 335, 437–459.

Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T., & Zilles, K. (2001). Human primary auditory cortex: Cytoarchitectonic subdivisions and mapping into a spatial reference system. *NeuroImage*, 13, 684–701.

Morosan, P., Rademacher, J., Palomero-Gallagher, N., & Zilles, K. (2005a). Anatomical organization of the human auditory cortex: Cytoarchitecture and transmitter receptors. In P.Heil, E. König, & E. Budinger (Eds.), *The auditory cortex: Towards a synthesis of human and animal research* (pp. 27–50). Mahwah, NJ: Lawrence Erlbaum.

Morosan, P., Schleicher, A., Amunts, K., & Zilles, K. (2005b). Multimodal architectonic mapping of human superior temporal gyrus. *Anatomy and Embryology*, 210, 401–406.

Musiek, F. E. & Reeves, A. G. (1990). Asymmetries of the auditory areas of the cerebrum. *Journal of the American Academy of Audiology*, 1, 240–245.

Nakahara, H., Yamada, S., Mizutani, T., & Murayama, S. (2000). Identification of the primary auditory field in archival human brain tissue via immunocytochemistry of parvalbumin. *Neuroscience Letters*, 286, 29–32.

Nissl, F. (1894). Über die sogenannten Granula der Nervenzellen. *Neurologisches Zentralblatt*, 13, 676–685.

Ong, W. Y., & Garey, L. J. (1990). Neuronal architecture of the human temporal cortex. *Anatomy and Embryology*, 181, 351–364.

Parker, G. J., Luzzi, S., Alexander, D. C., Wheeler-Kingshott, C. A., Ciccarelli, O., & Lambon Ralph, M. A. (2005). Lateralization of ventral and dorsal auditory-language pathways in the human brain. *NeuroImage*, 24, 656–666.

Penhune, V. B., Zatorre, R. J., MacDonald, J. D., & Evans, A. C. (1996). Interhemispheric anatomical differences in human primary auditory cortex: Probabilistic mapping and volume measurement from magnetic resonance scans. *Cerebral Cortex*, 6, 661–672.

Pfeifer, R. A. (1920). Myelogenetisch-anatomische Untersuchungen über das kortikale Ende der Hörleitung. Leipzig: Teubner.

Pfeifer, R. A. (1936). Pathologie der Hörstrahlung und der corticalen Hörsphäre. In O.Bunke & O. Foerster (Eds.), *Handbuch der Neurologie* (pp. 533–626). Berlin: Springer.

Rademacher, J., Caviness, V. S., Jr., Steinmetz, H., & Galaburda, A. M. (1993). Topographical variation of the human primary cortices: Implications for neuroimaging, brain mapping, and neurobiology. *Cerebral Cortex*, 3, 313–329.

Rademacher, J., Morosan, P., Schormann, T., Schleicher, A., Werner, C., Freund, H. J., et al. (2001). Probabilistic mapping and volume measurement of human primary auditory cortex 9. *NeuroImage*, 13, 669–683.

Rademacher, J., Burgel, U., & Zilles, K. (2002). Stereotaxic localization, intersubject variability, and interhemispheric differences of the human auditory thalamocortical system. *NeuroImage*, 17, 142–160.

Rivier, F., & Clarke, S. (1997). Cytochrome oxidase, acetylcholinesterase, and NADPH-diaphorase staining in human supratemporal and insular cortex: Evidence for multiple auditory areas. *NeuroImage*, 6, 288–304.

Rojas, D. C., Teale, P., Sheeder, J., Simon, J., & Reite, M. (1997). Sex-specific expression of Heschl's gyrus functional and structural abnormalities in paranoid schizophrenia. *American Journal of Psychiatry*, 154, 1655–1662.

Romani, G. L., Williamson, S. J., & Kaufman, L. (1982). Tonotopic organization of the human auditory cortex. *Science*, 216, 1339–1340.

Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., & Rauschecker, J. P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nature Neuroscience*, 2, 1131–1136.

Rubens, A. B., Mahowald, M. W., & Hutton, J. T. (1976). Asymmetry of the lateral (sylvian) fissures in man. *Neurology*, 26, 620–624.

Sacco, C. B., Tardif, E., Genoud, C., Probst, A., Tolnay, M., Janzer, R. C., et al. (2009). GABA receptor subunits in human auditory cortex in normal and stroke cases. *Acta Neurobiologiae Experimentalis (Warsaw)*, 69, 469–493.

Sarkissov, S. A., Filimonoff, I. N., Kononowa, I. P., Preobrazeanskaja, N. S., & Kukuewa, L. A. (1955). *Atlas of the cytoarchitectonics of the human cerebral cortex (in Russian)*. Moscow: Medgiz.

Schlaug, G., Jancke, L., Huang, Y., & Steinmetz, H. (1995). In vivo evidence of structural brain asymmetry in musicians. *Science*, 267, 699–701.

Schleicher, A., Amunts, K., Geyer, S., Morosan, P., & Zilles, K. (1999). Observer-independent method for microstructural parcellation of cerebral cortex: A quantitative approach to cytoarchitectonics. *NeuroImage*, 9, 165–177.

Schleicher, A., Palomero-Gallagher, N., Morosan, P., Eickhoff, S. B., Kowalski, T., de Vos, K. et al. (2005). Quantitative architectural analysis: A new approach to cortical mapping. *Anatomy and Embryology*, 210, 373–386.

Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., & Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nature Neuroscience*, 5, 688–694.

Schneider, P., Sluming, V., Roberts, N., Bleeck, S., & Rupp, A. (2005). Structural, functional, and perceptual differences in Heschl's gyrus and musical instrument preference. *Annals of the New York Academy of Sciences*, 1060, 387–394.

Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*, 26, 100–107.

Seldon, H. L. (1981). Structure of human auditory cortex. II. Axon distributions and morphological correlates of speech perception. *Brain Research*, 229, 295–310.

Smiley, J. F., Hackett, T. A., Ulbert, I., Karmas, G., Lakatos, P., Javitt, D. C., et al. (2007). Multisensory convergence in auditory cortex, I. Cortical connections of the caudal superior temporal plane in macaque monkeys. *Journal of Comparative Neurology*, 502, 894–923.

Spierer, L., Bellmann-Thiran, A., Maeder, P., Murray, M. M., & Clarke, S. (2009). Hemispheric competence for auditory spatial representation. *Brain*, 132, 1953–1966.

Steinmetz, H. (1996). Structure, functional and cerebral asymmetry: In vivo morphometry of the planum temporale. *Neuroscience and Biobehavioral Reviews*, 20, 587–591.

Steinmetz, H., Rademacher, J., Jancke, L., Huang, Y. X., Thron, A., & Zilles, K. (1990). Total surface of temporoparietal intrasylvian cortex: Diverging left-right asymmetries. *Brain and Language*, 39, 357–372.

Sweet, R. A., Dorph-Petersen, K. A., & Lewis, D. A. (2005). Mapping auditory core, lateral belt, and parabelt cortices in the human superior temporal gyrus. *Journal of Comparative Neurology*, 491, 270–289.

Talairach, J., & Tournoux, P. (1988). *Coplanar stereotaxic atlas of the human brain*. Stuttgart: Thieme.

Talavage, T. M., Sereno, M. I., Melcher, J. R., Ledden, P. J., Rosen, B. R., & Dale, A. M. (2004). Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. *Journal of Neurophysiology*, 91, 1282–1296.

Tardif, E., & Clarke, S. (2001). Intrinsic connectivity of human auditory areas: A tracing study with DiI. *European Journal of Neuroscience*, 13, 1045–1050.

Tardif, E., & Clarke, S. (2002). Commissural connections of human superior colliculus. *Neuroscience*, 111, 363–372.

Tardif, E., Chiry, O., Probst, A., Magistretti, P. J., & Clarke, S. (2003). Patterns of calcium-binding proteins in human inferior colliculus: Identification of subdivisions and evidence for putative parallel systems. *Neuroscience*, 116, 1111–1121.

Tardif, E., Delacuisine, B., Probst, A., & Clarke, S. (2005). Intrinsic connectivity of human superior colliculus. *Experimental Brain Research*, 166, 316–324.

Tardif, E., Probst, A., & Clarke, S. (2007). Laminar specificity of intrinsic connections in Broca's area. *Cerebral Cortex*, 17(12), 2949–2960

Upadhyay, J., Ducros, M., Knaus, T. A., Lindgren, K. A., Silver, A., Tager-Flusberg, H., et al. (2007). Function and connectivity in human primary auditory cortex: A combined fMRI and DTI study at 3 Tesla. *Cerebral Cortex*, 17, 2420–2432.

Van Buren, J. M., & C. Borke (1972). *Variations and connections of the human thalamus*. Berlin: Springer-Verlag.

Van der Zwaag, W., Gentile, G., Gruetter, R., Spierer, L., & Clarke, S. (2011). Where sound position influences sound object representations: A 7T fMRI study. *NeuroImage*, 54, 1803–1811.

Viceic, D., Fornari, E., Thiran, J. P., Maeder, P. P., Meuli, R., Adriani, M., et al. (2006). Human auditory belt areas specialized in sound recognition: A functional magnetic resonance imaging study. *NeuroReport*, 17, 1659–1662.

Vogt, C., & Vogt, O. (1919). Allgemeinere Ergebnisse unserer Hirnforschung. *Journal für Psychologie und Neurologie*, 25, 292–398.

von Economo, C., & Horn, L. (1930). Über Windungsrelief, Maße und Rindenarchitektonik der Supratemporalfläche, ihre individuellen und ihre Seitenunterschiede. *Zeitschrift für Neurologie und Psychiatrie*, 130, 678–757.

von Economo, C., & Koskinas, G. N. (1925). *Die Cytoarchitektonik der Grosshirnrinde des erwachsenen Menschen*. Berlin: Springer.

Wallace, M. N., Johnston, P. W., & Palmer, A. R. (2002). Histochemical identification of cortical areas in the auditory region of the human brain. *Experimental Brain Research*, 143, 499–508.

Warrier, C., Wong, P., Penhune, V., Zatorre, R., Parrish, T., Abrams, D., et al. (2009). Relating structure to function: Heschl's gyrus and acoustic processing. *Journal of Neuroscience*, 29, 61–69.

Weigert, C. (1882). Über eine neue Untersuchungsmethode des Centralnervensystems. *Zentralblatt für die medizinischen Wissenschaften*, 20, 753–774.

Westbury, C. F., Zatorre, R. J., & Evans, A. C. (1999). Quantifying variability in the planum temporale: A probability map. *Cerebral Cortex*, 9, 392–405.

Wiesendanger, E., Clarke, S., Kraftsik, R., & Tardif, E. (2004). Topography of cortico-striatal connections in man: Anatomical evidence for parallel organization. *European Journal of Neuroscience*, 20, 1915–1922.

Yamamoto, T., Williamson, S. J., Kaufman, L., Nicholson, C., & Llinas, R. (1988). Magnetic localization of neuronal activity in the human brain. *Proceedings of the National Academy of Sciences of the U S A*, 85, 8732–8736.

Zilles, K., Dabringhaus, A., Geyer, S., Amunts, K., Qu, M., Schleicher, A., et al. (1996). Structural asymmetries in the human forebrain and the forebrain of non-human primates and rats. *Neuroscience and Biobehavioral Reviews*, 20, 593–605.

Zilles, K., Palomero-Gallagher, N., Grefkes, C., Scheperjans, F., Boy, C., Amunts, K., et al. (2002a). Architectonics of the human cerebral cortex and transmitter receptor fingerprints: Reconcilingfunctionalneuroanatomyandneurochemistry.*EuropeanNeuropsychopharmacology*, 12, 587–599.

Zilles, K., Schleicher, A., Palomero-Gallagher, N., & Amunts, K. (2002b). Quantitative analysis of cyto- and receptor architecture of the human brain. In J. C.Mazziotta & A. Toga (Eds.), *Brain mapping: The methods* (pp. 573–602). Amsterdam: Elsevier.

Chapter 3
Invasive Research Methods

Matthew A. Howard III, Kirill V. Nourski, and John F. Brugge

Abbreviations

AEP	averaged evoked potential
CT	computed tomography
ECoG	electrocorticography
EEG	electroencephalography
ERBP	event-related band power
fMRI	functional magnetic resonance imaging
HDE	hybrid depth electrode
HG	Heschl's gyrus
IFG	inferior frontal gyrus
LFP	local field potential
MEG	magnetoencephalography
MRI	magnetic resonance imaging
PET	positron emission tomography
STG	superior temporal gyrus
vPFC	ventral prefrontal cortex

M.A. Howard III
Department of Neurosurgery, University of Iowa, 200 Hawkins Drive, 1823 JPP,
Iowa City, IA 52242, USA
e-mail: matthew-howard@uiowa.edu

K.V. Nourski (✉)
Department of Neurosurgery, University of Iowa, 200 Hawkins Drive, 1815 JCP,
Iowa City, IA 52242, USA
e-mail: kirill-nourski@uiowa.edu

J.F. Brugge
Department of Neurosurgery, University of Iowa, 200 Hawkins Dr. 1624 JCP,
Iowa City, IA 52242, USA
e-mail: john-brugge@uiowa.edu

D. Poeppel et al. (eds.), *The Human Auditory Cortex*, Springer Handbook
of Auditory Research 43, DOI 10.1007/978-1-4614-2314-0_3,

3.1 Introduction

Auditory cortex, in the classic sense of the term, is taken to be the cluster of anatomically and physiologically distinct areas of temporal neocortex that are uniquely and reciprocally connected with one another and with the medial geniculate body and related thalamic nuclear groups. In humans, as many as seven or eight anatomically distinct auditory cortical fields have been identified on the supratemporal plane and posterolateral superior temporal gyrus (STG) (see Clarke and Morosan, Chapter 2). Lying outside of the classical auditory cortical fields of humans are areas of the middle and inferior temporal gyri and of the anterior polar region of the STG, all of which are considered involved in speech and language processing (see Giraud and Poeppel, Chapter 9). Reciprocal connections between temporal auditory cortical fields and auditory-related areas of frontal and parietal lobes are pathways underlying higher-level auditory and auditory–visual processing including speech perception, goal-directed motor action, and feedback critical for the modulation of voicing (Romanski, 2004; Cohen et al., 2009).

Most research on auditory cortex has been, and continues to be, performed in experimental animals, including nonhuman primates, using invasive physiological and anatomical methods. These invasive techniques are best suited, and in many instances uniquely suited, to address fundamental questions about the functional organization of auditory and auditory-related cortex. The results of systematic physiological and anatomical studies using these approaches in monkeys have given rise to a working model of hierarchical and serial-parallel processing of acoustic information within the auditory forebrain (Kaas & Hackett, 2005). Because certain features of the auditory forebrain are shared between humans and nonhuman primates, this model has become an attractive starting point for studying its functional organization in human (Hackett, 2003, 2007, 2008; Rauschecker & Scott, 2009). Applying this model to humans, however, should be exercised with some constraint as there are more than 200 living species of primates, including humans, each having evolved distinct auditory–vocal specializations within its respective ecological niche (Preuss, 1995). Indeed, there may not even be an appropriate single "primate model" of auditory cortical organization to apply to humans, especially where speech, language, and other higher-level cognitive processes are concerned. Understanding the mechanisms that underlie these processes requires research performed on human subjects.

In recent years a wide range of experimental methods has become available to advance the understanding of the structure and function of human auditory cortex. Among the noninvasive approaches are electroencephalography (EEG) and magnetoencephalography (MEG), which record cortical activity at a distance using electrodes glued to the scalp or from sensors distributed around the head (see Alain & Winkler, Chapter 4, and Nagarajan, Chapter 5), and brain-imaging methods based on changes in cerebral blood flow, including positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) (see Talavage and Johnsrude, Chapter 6). Invasive approaches involve electrophysiological recording from and electrically stimulating or locally cooling cortex directly, usually in neurosurgical patients undergoing diagnosis and treatment of medically intractable epilepsy.

This research is performed either in the operating room (acute experiments) or in a specially equipped neurosurgical ward that allows for long-term clinical electrocorticography (ECoG) and video monitoring of patients with chronically implanted electrodes (chronic experiments). Each of these noninvasive and invasive approaches has its limitations, but when used in complementary ways can yield new information not obtainable by any single approach used alone.

ECoG refers to recording of electrical activity directly from the brain surface or from deep brain structures. When an acoustic stimulus is presented, a local field potential (LFP) may arise in an auditory or auditory-related area reflecting activity evoked by that stimulus in an ensemble of neurons in the vicinity of the recording electrode. A variant of this approach allows for recording from single neurons or neuronal clusters (Engel et al., 2005). Focal electrical stimulation, often performed in the same subject and applied to the same electrodes through which the ECoG is obtained, aims to map cortical sites critical for hearing, speech, and language by creating a reversible "functional lesion" and thereby temporarily disrupting cortical processing around the site of stimulation (Boatman, 2004; Sinai et al., 2005). Focal cooling also creates a functional lesion but by reversibly blocking synaptic transmission in a small cortical area beneath a cooling probe (Bakken et al., 2003). Focal electrical stimulation and electrophysiological recording may be used together to trace functional connections within and between cortical fields (Liegeois-Chauvel et al., 1991; Brugge et al., 2003, 2005; Greenlee et al., 2004, 2007; Matsumoto et al., 2004, 2007). Invasive brain research is opportunistic in nature in that it takes advantage of patients' willingness to participate during surgical procedures usually performed for accurate localization of a seizure focus.

This chapter presents a brief historic overview of intracranial studies of auditory cortex in humans followed by a description of intracranial methods currently employed in recording from, stimulating, and deactivating auditory cortex of human subjects. These experimental approaches are designed to address questions of the locations, boundaries, and interconnections of the multiple auditory fields that make up human cortex, and how each of these fields contributes to processing of auditory information.

3.2 Brief Historic Overview

Progress in invasive human brain research has paralleled advances in the field of functional neurosurgery, electronic engineering, computer technology, and signal processing. Although technical aspects of invasive human brain research have changed markedly over the years, the importance of a multidisciplinary research team, pioneered by neurosurgeon Wilder Penfield, has remained. Today, research of this kind draws heavily on the disciplines of anatomy, physiology, psychophysics, neuropsychology, radiology, theoretical modeling, statistics, acoustics, signal processing, electronic engineering, and computer programming.

In 1934, Penfield founded the Montreal Neurologic Institute, where neurosurgical care and human brain research were first seamlessly integrated and where many landmark scientific studies were performed (Penfield & Rasmussen, 1950; Penfield

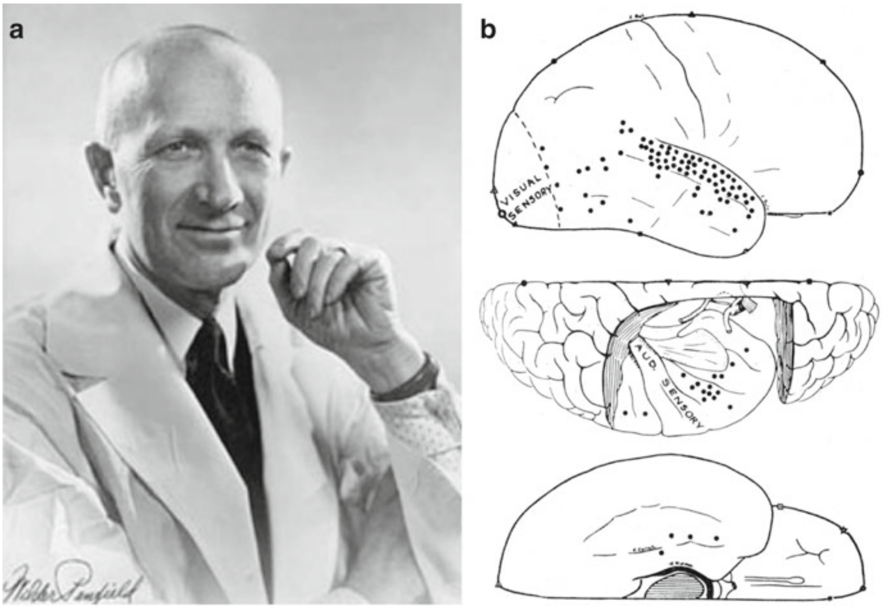


Fig. 3.1 (**a**) Wilder Penfield, pioneer neurosurgeon and founder of the Montreal Neurological Institute. (**b**) Sites on human temporal lobe from which electrical stimulation elicited experiential responses. Top: Lateral surface. Middle: Supratemporal plane; HG labeled AUD. SENSORY. Bottom: Inferior surface. (Adapted from Penfield & Perot, 1963.)

& Perot, 1963). With the awake patient's brain exposed, Penfield and his team carefully recorded bodily movements and verbal reports of sensations evoked by an electrical stimulus delivered to a cortical site through a hand-held stimulator. The anatomical location of each stimulated cortical locus was documented on a high-resolution intraoperative photograph of the brain surface. The results provided the first direct, systematic evidence of how human cerebral cortex is functionally organized. Auditory sensations aroused by cortical electrical stimulation were confined to sites on the lateral STG, or on the exposed supratemporal plane (Fig. 3.1). Crude auditory sensations (e.g., buzzing) were evoked by stimuli applied to sites deep within the Sylvian fissure, in what would now likely be considered the auditory core or belt fields. Complex auditory experiential hallucinations, on the other hand, were typically evoked by stimuli applied along the exposed lateral surface of the STG, though it is now thought that many of these latter sensations were the result of activating distant brain sites, including the limbic system (Gloor et al., 1982; cf. Moriarity et al., 2001).

Shortly after electronic recording instrumentation became available it was established that the brain itself was the source of the EEG and, further, that an acoustic stimulus could arouse an electrical event visible in the raw EEG trace (Davis, 1939). Another breakthrough came some 20 years later with the development of the laboratory computer and its use in extracting small acoustically evoked neural activity from background noise by signal averaging (Geisler et al., 1958). This opened the door for systematic study of the averaged auditory evoked potential

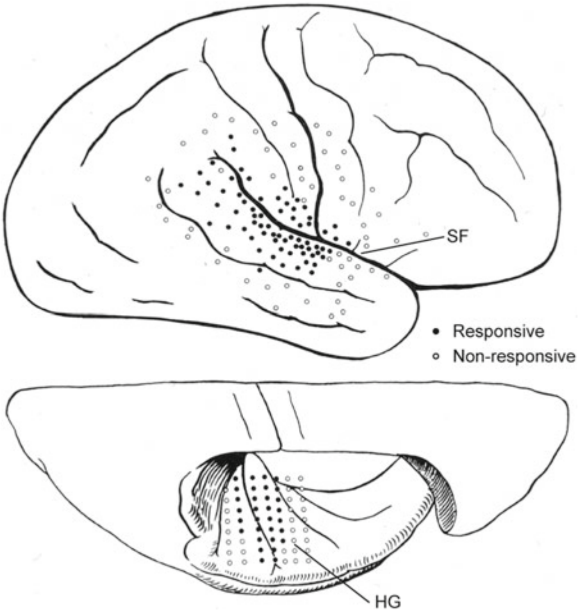


Fig. 3.2 Composite diagram showing cortical regions on the lateral brain surface (upper panel) and supratemporal plane (lower panel) explored with recording electrodes in acute experiments on 19 epilepsy surgery patients. Filled circles, sites excited by acoustic stimulation; open circles, no response to acoustic stimulation. HG, anterior transverse (Heschl's) gyrus; SF, Sylvian fissure. (Modified from Celesia, 1976.)

(AEP), and it was not long before this technology was introduced into the operating room where recordings were made through electrodes placed on different regions of the STG and supratemporal plane while sounds were presented to the patient during wakefulness and sleep and under general anesthesia (Fig. 3.2) (Celesia & Puletti, 1969; Celesia, 1976).

Much of what is now known about the functional organization of auditory cortex is derived from data obtained from single-unit studies performed in experimental animals using rigid metal or glass microelectrodes. Adapting this method for use in human subjects is particularly challenging because the potential for tissue damage represents an unacceptable human research risk under most circumstances. This risk is eliminated, however, by confining recording sites to regions of the temporal lobe that will be subsequently resected for clinical reasons. Experiments using this approach have generated important and unique data on the functional properties of auditory neurons within the cortex of the anterior STG and middle temporal gyrus (Creutzfeldt & Ojemann, 1989; Creutzfeldt et al., 1989a, b).

Technical advances in electrode fabrication have led to the development of hybrid depth electrodes (HDEs), which allow recording in human cortex of both LFPs from neuronal assemblies and action potentials from single neurons or neuronal clusters at multiple sites in auditory cortex deep within the supratemporal plane in awake patients over sustained periods of time (Howard et al., 1996b; Fried et al., 1999). HDEs are modified clinical depth electrodes and hence pose no additional surgical risk to the patient. With advances in electrode design and fabrication have come advances in computerized data acquisition, storage, and management systems capable of handling data obtained simultaneously from hundreds of recording sites. This in turn has been accompanied by innovative signal processing strategies.

3.3 Contemporary Research

3.3.1 Research Subjects

Invasive studies of auditory cortex in humans are carried out in neurosurgical patients, the vast majority of whom are being treated for medically intractable epilepsy. Epilepsy is a common neurological disorder that affects approximately 50 million people worldwide (World Health Organization, 2005). These individuals are at risk of losing consciousness suddenly and without warning and thus often are unable to operate a motor vehicle or hold a job that requires sustained vigilance and attention. There is also evidence that persistent seizure activity may lead to structural brain damage (Bonilha et al., 2006; Bernhardt et al., 2009). It is estimated that in the United States alone, more than 400,000 individuals with epilepsy continue to have seizures despite receiving appropriate medical treatment (Engel, 2001). A subset of this medically refractory patient population can be treated effectively with surgery that removes brain tissue that is the source of the seizure activity.

Candidates for resection surgery must fulfill three criteria. First, they must have failed to respond to medical management. Second, their quality of life would be markedly enhanced by achieving a seizure-free surgical outcome. The ideal surgical candidate is a young person whose educational and vocational opportunities promise to be enhanced substantially by eliminating seizures. These patients also are ideal subjects for invasive brain research. Third, their seizure focus must be localized to a circumscribed portion of the brain that can be safely removed surgically. Making this latter determination is the most challenging aspect of the presurgical evaluation process, which includes EEG recording, anatomical magnetic resonance imaging (MRI), and formal neuropsychological testing. The patient's history and test data are discussed at a multidisciplinary epilepsy surgery conference where a consensus is sought regarding whether the patient is a candidate for surgery, and, if so, the type of operation that should be performed. Surgical patients then wishing to participate in research are given a detailed explanation of the planned research protocols and provide informed consent before becoming a "research subject" and undergoing additional "research only" preoperative testing.

Once a patient has agreed to participate as a subject in research, structural (and in some instances functional) MRI scans are obtained. These images provide the subject's preoperative anatomical template. The locations of experimental recording sites are subsequently superimposed on the preoperative image set. Subjects involved in auditory research may also undergo a preoperative hearing evaluation to objectively measure possible hearing deficits. Acoustic stimuli used during experiments are typically generated digitally and may be delivered in the open field or through earphones. The acoustical properties of a stimulus are difficult to control in the open field, especially in a clinical environment that may have many reflecting surfaces and high levels of ambient background sound. Over-the-ear headphones are impractical to use because the head bandage does not allow for a good acoustical seal. Insert earphones, on the other hand, may be integrated into custom fitted ear molds of the kind commonly worn by hearing aid users. Ear molds created for each subject conform to the anatomy of a subject's external ears, thereby providing an acoustic seal that assures accurate sound delivery while attenuating unwanted ambient noise; their snug fit also resists dislodgement by head movement. Importantly for chronic experiments, these earphones can be removed repeatedly and reliably reinserted.

3.3.2 Acute Experiments

In a typical epilepsy center, approximately half of the patients who are deemed to be candidates for resection surgery will undergo the procedure without additional preoperative diagnostic testing. In these cases experiments are performed in the operating room, when the subject is awake and alert and able to communicate with the research team and follow instructions. To gather additional information about the location of the seizure focus, two 30-minute ECoG recordings are obtained directly from the brain using multicontact grid and strip electrodes placed over the lateral surface of the exposed lateral and inferior surfaces of the temporal lobe. Experimentation is permitted during these intraoperative ECoG recording sessions. In acute experiments, recording and stimulating devices of various configurations are permitted for research purposes (Fig. 3.3). These include densely spaced multicontact recording and stimulating arrays, penetrating microelectrodes, stimulating probes, and local cooling devices (Fig. 3.4). With a typically wide exposure of the brain, it is safe to place and reposition these devices directly on the brain surface without injuring underlying tissue. The primary limitations of intraoperative experimentation relate to the time available for conducting experiments and the types of research tasks that subjects are capable of performing. The 30-minute time windows of opportunity are sufficiently long to perform experiments successfully if they are performed efficiently. If technical problems arise (e.g., electric power line noise), there is little time available to solve them. This places a premium on careful preoperative research planning and equipment testing. Subjects are typically in a supine position with a cushion under one shoulder. Surgical drapes are arranged to form a sterile barrier while at the same time allowing the patient to have a clear view of the anesthesiologist. Because most of the patient's body is covered in surgical drapes, and only minimal movement is allowed, the manual operations the patient can perform are usually limited to such simple tasks as verbalizing or button pushing. Finally, there is a high level of ambient noise in the operating room, which makes this a challenging environment for auditory experimentation, insert earphones notwithstanding.

3.3.3 Chronic Experiments

In some cases the results of the preoperative evaluation strongly suggest that although the patient is a candidate for resection surgery, there is some residual uncertainty about the location of the seizure focus. For these cases additional diagnostic

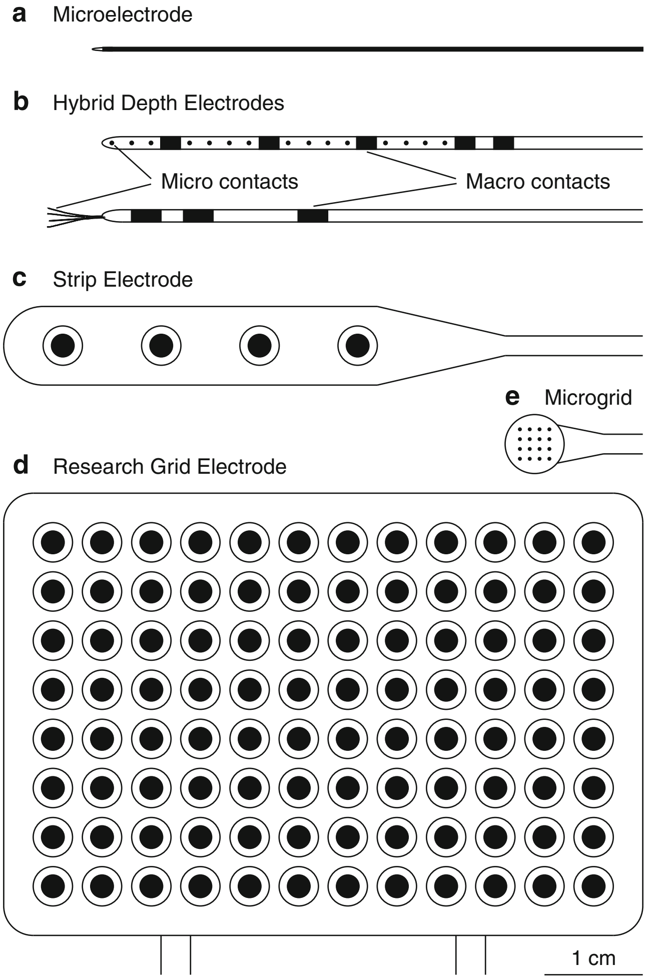


Fig. 3.3 Schematic diagram of types of electrodes used in invasive studies of human auditory cortex. (**a**) Microelectrode. Metal (usually tungsten) microelectrode, insulated except for the tip, capable of recording single neurons or neuronal clusters (multiunit activity) and used in acute studies of human auditory cortex. (**b**) Hybrid depth electrodes. Modified clinical electrodes having large (macro) contacts capable of recording field potentials, and small (micro) contacts capable of recording both field potentials and action potentials. Two types of HDEs are shown. Top: Micro contacts are cut ends of microwires conforming to the shaft of the electrode. Bottom: Micro contacts are micro wires that are extruded from the end of the electrode after implantation. (**c**) Strip electrode. Clinical electrode with macro contacts used both in acute and chronic studies. (**d**) Research grid electrode. Clinical grid electrode modified such that macro contacts in the array are separated by 4–5 mm (rather than 10 mm) and used primarily in chronic studies. (**e**) Microgrid. Surface grid electrodes with microwire contacts having 1 mm separation

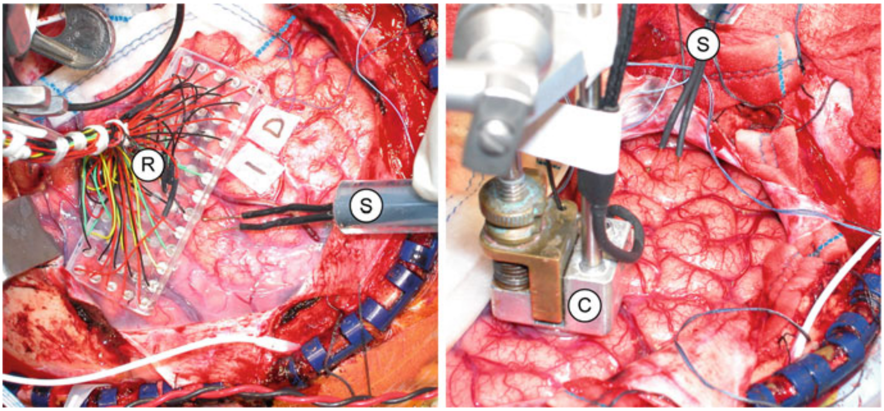


Fig. 3.4 Acute experiments. Photographs taken in the operating room showing a recording grid (R), hand-held bipolar electrical stimulator (S), and cooling probe (C) in direct contact with brain surface

testing is needed, which may include ECoG recording from electrode arrays chronically implanted directly on the pial surface or within cortical tissue, or both. Experiments are subsequently carried out in a specially equipped hospital suite while the patient is recovering from implant surgery and undergoing longer-term clinical video and ECoG monitoring. The proportion of patients undergoing invasive ECoG monitoring before resection surgery varies from one clinical research center to another, and the devices and techniques used to perform this monitoring vary as well. The electrodes, electronic instrumentation, and experimental protocols associated with chronic recording may also differ substantially from those of acute experimentation. When used in complementary ways, however, acute and chronic recording together provide important information about auditory cortical organization not gained by relying on one approach alone.

All chronic invasive research on auditory cortex of humans is performed using standard, or custom modified, clinical ECoG electrodes (see Fig. 3.3). These devices fall into two broad categories: depth electrodes and surface arrays. Patients undergo implantation of recording electrodes directly in and on the brain in the vicinity of the suspected seizure focus (Fig. 3.5). Implantation is performed under general anesthesia, and usually no experiments are conducted during this stage.

3.3.3.1 Depth Electrodes

Depth electrodes are designed for clinical ECoG recordings from brain sites deep beneath the cortical surface. They are thin, flexible, silicon-based cylinders, typically 1–2 mm in cross-sectional diameter with low impedance, circumferential platinum contacts positioned along the electrode shaft. For experimental purposes a clinical depth electrode may be modified to create a “hybrid,” with additional contacts added for higher spatial resolution and capable of recording both LFPs and single

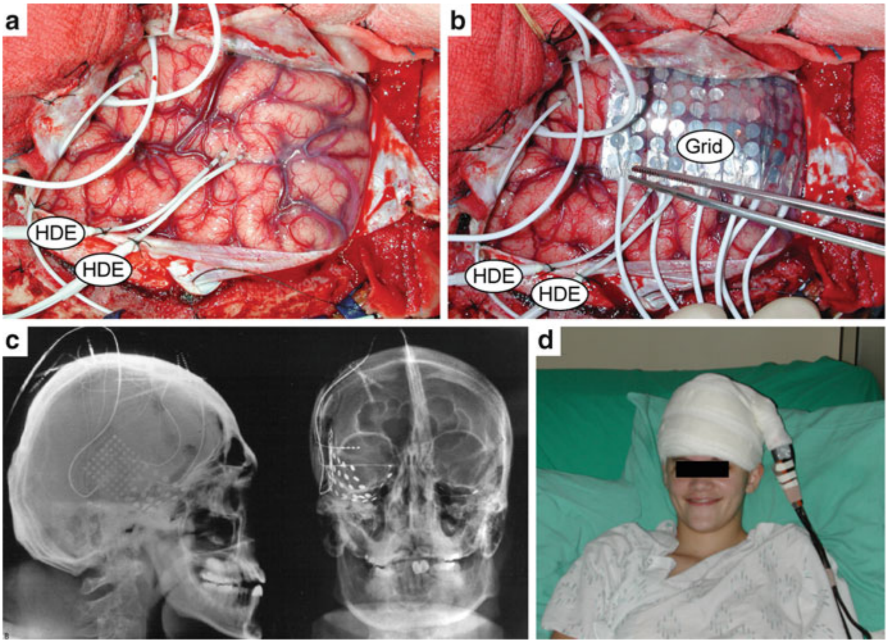


Fig. 3.5 Chronic invasive ECoG monitoring. Top: Photographs taken during electrode implantation showing the sites of entry of two HDEs (**a**) and the placement of a research grid array (**b**). (**c**) Lateral and frontal postimplantation head X-ray images showing a 64-channel grid array over perisylvian cortex, a frontal grid, and two HDEs, one in HG and the other in the amygdala. (**d**) Postimplantation photograph of a research subject

neurons or neuronal clusters (see Fig. 3.3b). One type of HDE has insulated platinum–iridium microwires (25–50 μm diameter) inserted into the lumen of the clinical electrode with cut ends brought to the surface of the shaft at regular (~2 mm) intervals between the clinical ECoG contacts (Howard et al., 1996b). This linear array of 14–16 microcontacts is particularly well suited to mapping response properties of auditory fields on Heschl's gyrus (HG), as described later in this chapter. A different type of HDE also employs insulated microwires that, once the electrode is in place, can be extruded beyond the distal tip of the shaft (Fried et al., 1999). This approach may be more suitable for isolating single neurons or neuronal clusters because recordings can be made some distance from the electrode shaft, but it has the disadvantage that the locations of recording sites are difficult to specify and later to identify. The requirement for microwires, however small in diameter, is one of the limiting factors in using HDEs as currently designed, as only a relatively small number can be fitted into the electrode shaft. In addition to carrying microwires, the HDE may also be equipped with a microdialysis probe within its lumen capable of sampling in situ neuroactive substances in the extracellular milieu (Fried et al., 1999).

A slotted cannula, temporarily inserted stereotactically into the brain parenchyma using standard or minimally modified neurosurgical techniques, serves as a

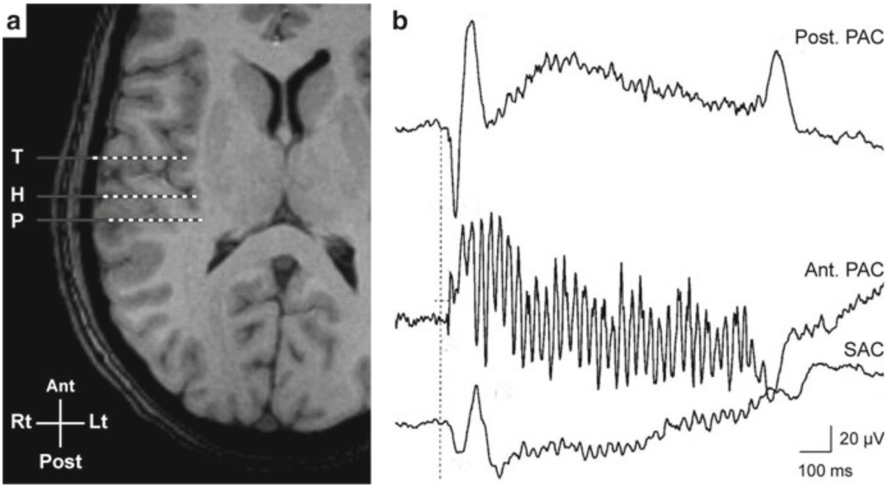


Fig. 3.6 Multiple depth electrodes implanted in the supratemporal plane. (**a**) Horizontal MR image through supratemporal plane of a human epilepsy surgery patient showing the trajectory (dashed lines) of three chronically implanted electrodes. (**b**) AEPs recorded from posterior (Post. PAC) and anterior (Ant. PAC) core cortex on Heschl's gyrus and from secondary auditory cortex (SAC). (Modified from Liégeois-Chauvel et al., 2004.)

guide to the implantation of the flexible HDE. Two broad categories of stereotactic systems are available for use: rigid frame and frameless systems. Because of the highly complex contour of the supratemporal plane in humans, including the orientation of the transverse gyri, along with the considerable intersubject variability in the overall structure of the STG, the choice of the frame system used impacts the way in which auditory cortex on the supratemporal plane is approached anatomically and, hence, how the research is carried out.

Rigid frame devices are firmly attached to the patient's skull with multiple fixation points before surgery. The frame serves as a fixed three-dimensional spatial reference system for target selection and as a platform to which an electrode insertion device is attached. In contemporary practice, the patient typically undergoes a brain imaging study (usually MRI or computerized tomography [CT]) with the frame in place. Cerebral angiography, which provides images of cerebral blood vessels, may also be performed. Using imaging data, the three-dimensional coordinates for each brain target are calculated within the frame-based spatial reference system. This method may be used to implant electrodes in auditory cortex by neurosurgeons trained in the Talairach school of stereotactic surgery. The Talairach coordinate system is one devised to describe, in three-dimensional space, the location of brain structures across individuals independent of their brain shape or size. With this approach, a patient's brain images are used to select targets and choose electrode trajectories. Multiple depth electrodes may be inserted through burr holes in the skull along straight, lateral-to-medial trajectories that are parallel to one another and perpendicular to the sagittal plane (Fig. 3.6). Liégeois-Chauvel and colleagues (Liégeois-Chauvel et al., 1991, 1994, 1999, 2004) have taken full advantage of this arrangement to record auditory evoked activity over a wide area of the supratemporal plane. Although this approach allows functional study of several areas of auditory cortex including those on the planum temporale, it is not the optimal approach to the auditory core and belt area(s) located on HG. In this approach, HG is oriented obliquely with respect to the trajectory of the electrodes. Each depth electrode, therefore, crosses cortical laminae obliquely as it traverses a restricted segment of the gyrus, and as a consequence the sampling of auditory cortex on HG is necessarily limited.

Frameless stereotactic methods were developed to allow surgeons access to brain sites with stereotactic precision without being encumbered by the mechanical constraints associated with rigid head frames. A variety of frameless systems have been developed over the years, but the systems used most often in contemporary neurosurgical practice are based on infrared spatial-localization technology. Before surgery, fiducial markers are placed at multiple locations around the patient's scalp. An anatomical MRI is obtained, which includes the fiducial markers, and from this image a three-dimensional brain space is created. In the operating room, the patient's head is secured in a stationary position while the three-dimensional image is used to guide the depth electrode to its intended target(s). Because there are no mechanical constraints associated with the frameless system, there are no physical restrictions on the selection of electrode trajectories. At the University of Iowa, for example, the flexibility of the frameless system has been exploited to develop an electrode implantation technique that results in placement of the entire shaft of a depth electrode within HG (Reddy et al., 2010). The electrode is introduced into the cortex at the anterolateral boundary of HG, and then gently inserted along the crest of the gyrus in a direction oblique to the cortical surface. Because the electrode is somewhat flexible, even with the insertion stylette in place, it usually stays within the gray matter of the gyrus without penetrating the pial surface and entering the overlying Sylvian fissure. When properly performed, this method results in the placement of a linear array of 18 or more recording contacts along the full length of HG (Fig. 3.7). This approach has made it possible to obtain a spatial pattern of recordings that reveal the transition from presumed core auditory cortex, to more anterolaterally positioned belt, or parabelt fields, all located within HG (Brugge et al., 2008, 2009; Nourski et al., 2009). The entry point of this electrode is anterior enough on the STG to allow a grid array to be implanted on auditory cortex posterior to it, thereby providing an opportunity to record simultaneously from auditory cortical fields of HG and posterolateral STG.

Whereas stable LFPs are routinely obtained from chronically implanted penetrating microelectrodes, recording from single neurons or neuronal clusters is a greater challenge. The problem of obtaining high-quality recording arises in large part from the reactive response of cortical tissue to the chronic implant. After implantation, a glial barrier forms around the electrode shaft, which over time effectively reduces the signal-to-noise ratio of the recordings (Pierce et al., 2009). Various approaches are being used in an attempt to rejuvenate microelectrode sites, aimed at increasing biocompatibility, reducing electrode impedance, and improving electrode interface properties (Johnson et al., 2004; Lempka et al., 2006).

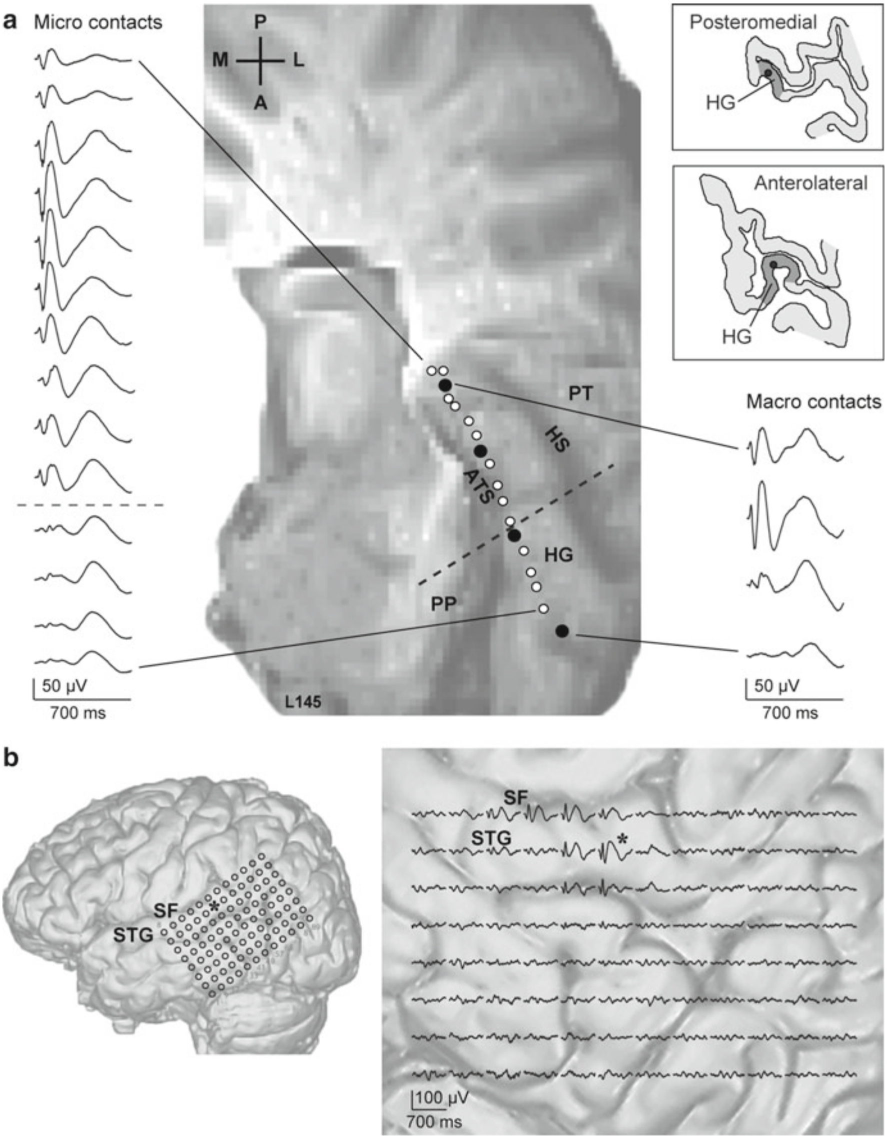


Fig. 3.7 Responses to click-train stimulation, recorded simultaneously from a HDE and a research grid electrode. (**a**) MR image of supratemporal plane showing the locations of micro and macro contacts on an HDE chronically implanted in Heschl's gyrus and click-evoked AEPs recorded at these locations. Insets: Line drawings of MRI cross sections showing the position of the electrode within posteromedial (core) and anterolateral (belt) cortex. Approximate boundary between core and belt indicated by dashed line. (**b**) Location of a 96-contact research grid on an MRI of the lateral brain surface (left), with AEPs superimposed (right). (Modified from Brugge et al., 2008.)

3.3.3.2 Surface Grid Electrodes

Research grid arrays, which are commonly rectangular in shape and consist of flat low impedance platinum–iridium discs within a flexible silastic membrane (see Fig. 3.3d), are placed on the pial surface of the brain under direct visual control. Strip electrodes, which are fabricated from the same materials but usually consist of only a single row of contacts, typically have a narrow configuration (see Fig. 3.3c). This makes them well suited for insertion into subdural spaces beyond the boundaries of the craniotomy. It is common practice, for example, to insert strip electrodes into the subdural space at the inferior boundary of a temporal craniotomy and gently slide the array under the temporal lobe to obtain coverage of the ventral surface of the temporal lobe, an area that cannot be visualized directly during surgery. Standard clinical grid and strip electrode are manufactured with 1 cm spacing between contacts. This spacing is adequate for purposes of obtaining clinical ECoG recordings, and many investigators have used these same electrode arrays to perform electrical stimulation (Boatman, 2004) and electrophysiological recording (Crone et al., 2001) experiments. When higher spatial resolution is called for in research studies, modified surface arrays with as little as 4 or 5 mm spacing between contacts are employed (Howard et al., 2000; Brugge et al., 2008; Flinker et al., 2011). Surface grid electrodes with an even finer grain (1 mm separation) designed for studies of human brain computer interface have been introduced to studies of the cortical representation of speech (Kellis et al., 2010; see Fig. 3.3e). With modern cable and connector technology it is feasible to implant a patient with more than 200 depth and surface recording contacts without increasing the surgical risks. As with any wire conductors, however, electrical noise is easily coupled to them, and reducing this kind of interference is a particular challenge in a clinical environment. Further, wires may break, especially when a subject experiences an epileptic seizure. Wireless systems for cochlear stimulation have long been used to help restore hearing in deaf individuals, and such wireless multichannel systems are under development for electrical brain stimulation through chronically implanted microarrays (Ghovanloo et al., 2004).

3.3.3.3 Anatomical Reconstruction

Studies of functional organization of auditory cortex, whether in humans or nonhuman experimental animals, require accurate anatomical reconstruction of each recording and stimulation site. In human studies each experimental subject commonly undergoes whole-brain MRI scanning before and after implant surgery. Pre- and postimplantation MRIs are then co-registered and coordinates for each electrode contact obtained from postimplantation MRI volumes are transferred to preimplantation MRIs. This approach may include reliance on stereotaxic coordinates as well as CT and angiographic imaging. The position of depth electrodes and location of each recording site is then reported in Talairach coordinates. A related approach utilizes lateral and anterior–posterior X-ray imaging along with skull landmarks for co-registration with the standardized Talairach coordinate system (Miller et al., 2007). There are several confounding factors faced in anatomical reconstruction of recording sites. The first is that the presence of metallic electrodes in the brain creates substantial artifact in the postimplant MR image. To circumvent this, a companion CT image may be used along with MRI. Second, the head of the subject is not in exactly the same position for pre- and postoperative MR imaging. Yet another factor, which applies mainly to the HDE reconstruction, is that implantation surgery with the introduction of grid arrays and HDEs displaces the cerebral hemisphere medially, with superficial brain tissue being distorted more than deeper structures. Accurate assignment of recording loci on preimplantation MR images may require adjustments based on careful comparison of surrounding brain structures between the pre- and postimplantation MRI volumes. Nonlinear compression causes the depiction of electrode trajectory and the spacing of contacts, when transferred to the preoperative images, to appear irregular, as seen for example in Figure 3.7, where the location of each contact is projected on to the surface of HG.

While the trajectory of a depth electrode may be depicted by transferring the recording loci to a rendering of the cortical surface (see Fig. 3.7), it is equally important to know for each contact where in the depth of the cortex recordings were made. This is accomplished by obtaining serial MR cross-sectional images containing each of the depth recording contacts. The optimal orientation of the MR volumes should show the cortical grey matter in cross section, which for HG would be oblique with respect to the standard coronal plane of the brain. With the HDE inserted along the long axis of HG, for example (Howard et al., 1996b), these cross sectional images are obtained roughly orthogonal to the trajectory of the electrode. Either MR images or outline drawings derived from them along with reconstruction of the electrode trajectory depict the position of each recording contact in detail (see Figs. 3.6 and 3.7).

3.3.3.4 Stimulation and Recording

After electrode implantation and surgical recovery, patients are transferred to a monitoring facility. Here the epileptologist uses information from video and ECoG recordings to formulate a hypothesis as to the site(s) of origin of the patient's seizures. Meanwhile the patient is awake and alert during many of the daytime hours while clinical monitoring is progressing, and, though tethered by a cable bundle to the EEG monitoring equipment, is usually capable of participating in a wide range of research protocols.

Initially the patient is maintained on antiepileptic medications, but is then slowly weaned from these drugs during the monitoring period. In most instances, patients are able to participate fully in research protocols 2–3 days after the implantation surgery is performed. A typical experimental session is scheduled to last approximately 2 hours, including the time needed to arrange specialized equipment for specific protocols. With the exception of a head dressing and a thick bundle of cables exiting the bandage (see Fig. 3.5d), there are no mechanical or environmental constraints on a patient's ability to participate in experimental protocols. Thus, along with verbal responses and button presses, patients usually have no difficulty operating a computer mouse or other mechanoelectric devices. Patients are able to move from the hospital bed to a chair and back without assistance, and usually have no difficulty sitting upright for a 2-hour research session without becoming fatigued. Many are capable of participating in two, and some in as many as three, experimental sessions per day.

There are several advantages of chronic over acute cortical recording and stimulation. First, the time constraints associated with acute experimentation are not associated with chronic work, hence substantially more physiological data can be gathered, and experiments can be replicated in the same subject over days. The latter consideration becomes important as the stationarity of neuronal response properties could be affected by a number of factors changing over time, including slight postoperative movement of the recording array, recovery from trauma associated with electrode placement, or insertion and withdrawal of antiepileptic medication. Second, by using both electrode grid arrays for surface recording and HDEs for depth recording in the same subject it is possible to obtain data simultaneously from auditory cortical areas within HG and from other auditory fields represented on the supratemporal plane and lateral surface of the brain. By doing so, functional differences observed across recording sites can be attributed to differences in auditory field representations rather than to differences between experimental subjects or experimental conditions, or both. Third, patients are able to perform relevant tasks during recording and stimulation sessions, which allows for studies of relationships between brain activity and the level of task performance (Jenison et al., 2011).

Finally, at the end of the clinical monitoring period and when research recording and stimulation are coming to a close, an opportunity often exists to perform one last, and very limited, experiment to test the effects of general anesthesia on a particular aspect of auditory cortical physiology. This session is carried out in the operating room just before removal of the electrodes and surgical resection. Recording begins just before induction of anesthesia and is continued through loss of consciousness.

Although the chronic experimental setting has enormous advantages, there are also some important limitations to this approach as well, which have to be taken into account in designing experiments and interpreting results. Perhaps the most important relates to the fact that electrode arrays can be placed only in cortical areas dictated by clinical criteria. With considerable intersubject variation in the structure of the STG and the locations and boundaries of auditory cortical fields in it (Rademacher et al., 1993; Leonard et al., 1998), even in cases where extensive arrays are implanted, there are many auditory areas simply not sampled. This is particularly true for cortex within sulci even though penetrating electrodes may be placed into the cortex of the supratemporal plane. One approach that promises to address this limitation is to combine, in the same subject and under the same stimulus conditions, chronic intracortical recording with fMRI, taking advantage of the former for obtaining highly localized physiological measures and the latter for obtaining a global view of cortical activity (Mukamel et al., 2005). Success with this approach will require a better understanding of the relationship between the fMRI signal and neural activity recorded with implanted electrodes (see Cariani and Micheyl, Chapter 13).

3.4 Experimental Paradigms

The use of invasive methods to study the auditory and auditory-related cortical areas in humans with modern technology has provided opportunities to ask questions related both to the fundamental organization of these parts of the brain and to cortical mechanisms of speech and language processing that may be beyond the realm of study in nonhuman animals.

The organizational framework that forms the foundation for the understanding of auditory cortex is constructed around the concept of multiple interconnected fields, differentiated from each other anatomically and physiologically and each contributing to processing acoustic information in its own way. Extensive studies in experimental animals have shown the existence of multiple auditory fields in temporal cortex. As many as a dozen or more have been demonstrated anatomically in monkeys (Hackett, 2003, 2007, 2008). These fields have been shown to differ in their anatomical locations, neuronal response properties and connectivity patterns. While it has been shown in postmortem tissue that humans may exhibit 7 or 8 such fields on the STG (see Clarke and Morosan, Chapter 2), it has not been possible to use many of the anatomical and physiological approaches that have made experimental work in animals so fruitful. The successes in functionally identifying auditory fields in humans by means of invasive approaches have come through the use of three methodologies: electrophysiological recording, electrical stimulation tract tracing, and creation of functional lesions through focal electrical stimulation.

3.4.1 Functional Mapping by Electrophysiological Recording

Based mainly on mapping studies in human subjects with chronically implanted electrodes, a small area on posteromedial HG has been demonstrated to exhibit response properties that are consistent with it being the primary and primary-like (core) auditory cortex (Liégeois-Chauvel et al., 1991; Howard et al., 1996a; 2000; Brugge et al., 2008, 2009). It differs in fundamental ways from the area around it on the supratemporal plane and the lateral surface of the STG (see Figs. 3.6 and 3.7). The responses recorded from posteromedial HG differ from those recorded from the posterolateral STG in the overall morphology of the polyphasic AEPs and in specific physiological response properties (Howard et al., 2000). Compared to the posteromedial HG, cortex on the posterolateral STG is characterized by a slower recovery from previous stimulation, a lower phase-locking capacity, and a greater sensitivity to general anesthesia (Howard et al., 2000; Brugge et al., 2008).

To date, little is known of the functional properties of the auditory core and even less about multiple fields that surround it. Advances in this research area will require the use of complex auditory stimuli, and subjects will need to be engaged in tasks related to attention and higher cognitive processes including those related to speech and language. Because human communication engages the other sensory systems as well, vision and touch need to be introduced into stimulus paradigms. These issues and related computational challenges are addressed extensively in later chapters.

3.4.1.1 Signal Processing

As described in the historical overview, the development of computers capable of averaging evoked field potential activity was one of the technical breakthroughs that allowed systematic and quantitative study of evoked field potentials recorded from human auditory cortex. At the time there was no other practical means of detecting the low-amplitude evoked voltage deflections that were obscured by ongoing background activity. Computerized averaging methods became an indispensible element in the armamentarium of researchers investigating the physiological properties of auditory cortex of humans and nonhuman mammals. As long as the evoked response is precisely time locked to the onset of an auditory stimulus, and if a sufficient number of stimulus presentations are given, the random background activity is reduced (and the signal-to-noise ratio is enhanced) through the averaging process. As with any method, simple signal averaging in the time domain has limitations. Perhaps the most significant of these is that the relatively low-frequency AEP does not capture non–phase-locked field potential activity, particularly oscillatory responses at relatively high frequencies (>70 Hz) (termed "high gamma" range).

The biological importance of non–phase-locked cortical activity was first established in experimental animal studies (Freeman, 1978; Gray et al., 1989; Engel et al., 1991). More recent experiments performed in nonhuman primates have provided additional information regarding the cellular mechanisms mediating these high-frequency responses within auditory cortex (e.g., Steinschneider et al., 2008). Crone and his colleagues at Johns Hopkins University have studied successfully this so-called "induced" activity in human auditory cortex by combining ECoG recording with signal processing methods that measured the spectral content of the stimulus-related brain activity (Crone et al., 2001, 2006). There is now convincing evidence from several laboratories indicating that non–phase-locked high-frequency activity recorded from human auditory cortex contains information about the acoustic stimulus not found in the AEP (Ray et al., 2008; Brugge et al., 2009; Edwards et al., 2009; Nourski et al., 2009).

Spectral analytic methods (e.g., fast Fourier transform, wavelet transform) can now be efficiently performed on field potential data using standard computers. Using these techniques, it is feasible to objectively measure stimulus-induced power changes—the so-called event-related band power (ERBP)—throughout the spectral range of the evoked response. Although the *absolute* ECoG power in the high-frequency range is very low compared to that in the low-frequency range, ERBP, which represents *proportional* changes in power after sensory stimulus presentation compared to a prestimulus baseline, can be much greater in the high-frequency range than in the low-frequency range.

The application of time–frequency ERBP analysis in studies of human auditory cortex is illustrated in Figure 3.8. Here, responses of core auditory cortex to a variety of acoustic stimuli are displayed as AEP waveforms and time–frequency ERBP plots. Trains of acoustic transients evoke frequency-following responses, evident at relatively low repetition rates, as well as increases in high-frequency ERBP (Fig. 3.8a; Brugge et al., 2009). Regular-interval noise, generated by introducing temporal

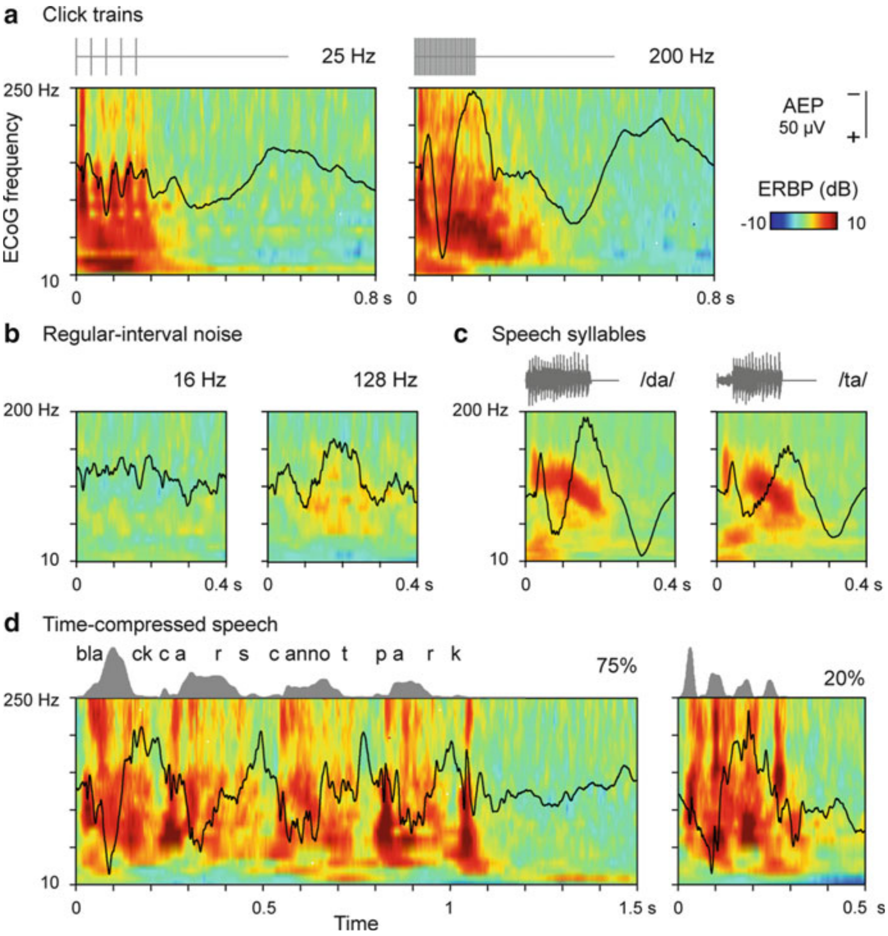


Fig. 3.8 ERBP analysis of intracranial recordings recorded from human core auditory cortex in response to different acoustic stimuli. (**a**) Click trains presented at 25 (left) and 200 (right) Hz (replotted from Brugge et al., 2009). (**b**) Transition from random to regular-interval noise, generated using delays corresponding to 16 (left) and 128 Hz (right) periodicity (replotted from Griffiths et al., 2010). (**c**) Speech syllables /da/ (left) and /ta/ (right). (**d**) Speech sentence "black cars cannot park," time-compressed to 75% (left) and 20% (right) of its original duration (replotted from Nourski et al., 2009). AEP waveforms are superimposed on the ERBP time–frequency plots. Stimulus schematics are shown in gray

regularity to a broadband noise stimulus, elicits an increase in ERBP when this temporal regularity is associated with a pitch percept, but not when the repetition rate is below pitch frequency range (Fig. 3.8b; Griffiths et al., 2010). When presented with speech utterances, patterns of ERBP within core auditory cortex represent voice onset time, that is, the time interval between consonant release and the onset of voicing (Fig. 5.8c; see also Steinschneider et al., 2005). Temporal envelope of a speech sentence can be tracked by cortical high gamma activity when the stimulus is moderately compressed in time (accelerated), as well as at greater degrees of compression that make the sentence unintelligible (Fig. 3.8d; Nourski et al., 2009). Spectral-based signal processing such as shown in these examples has evolved to become a standard analytical approach in modern invasive human auditory cortex research.

3.4.1.2 Coding of Stimulus Acoustic Features

Traditionally, studies of auditory cortex in humans and nonhumans have involved presenting a stimulus and recording the electrophysiological response of single neurons or ensembles of neurons. Implicit in this approach is that buried in the responses recorded is the information being transmitted to and through the cortex—the code for that particular stimulus or stimulus attribute. A number of coding mechanisms (e.g., rate, time, place) are generally agreed upon. Evidence for their presence is provided by analysis of physiological data and its relationship to behavior.

The frequency content of a sound is a strong identifier of the sound source, and becomes particularly important in human speech communication. Many auditory cortical neurons are responsive to a restricted range of stimulus frequencies, referred to as their frequency response areas. Such neurons are typically most sensitive to a narrow range of frequency, the center of which is referred to as the “best” or “characteristic” frequency. Frequency tuning has been considered one mechanism by which frequency is discriminated, and auditory cortex has been considered a place where requisite neurons are located. Single neurons in HG of human subjects recorded with implanted HDEs have been found that are extraordinarily narrowly tuned (“ultra sharp”), and their frequency selectivity may account for a listeners threshold of frequency discrimination as measured psychophysically (Bitterman et al., 2008). Tuning curves similar to those recorded in auditory cortex of laboratory animals have also been recorded in human HG, and their distribution has confirmed the presence of at least one tonotopic field in the human auditory core (Howard et al., 1996a).

The amplitude and frequency of natural sounds, including speech, vary over time, and the auditory system has evolved mechanisms for detecting amplitude and frequency modulations. For slowly varying amplitude-modulated stimuli, below about 50 Hz, auditory cortical neurons in monkey phase-lock strongly to the modulation envelope, and hence encode the modulation frequency “explicitly” in the temporal cadence of their discharge (Steinschneider et al., 1998; Lu et al., 2001). Modulation envelopes in running speech in this frequency range are associated with individual words, syllables, and phonemes (Rosen, 1992). Local field potentials recorded in the human auditory core by means of HDEs implanted in HG show locking to repeated transients over a frequency range similar to that of monkeys (Liégeois-Chauvel et al. 2004; Brugge et al. 2009; see Figs. 3.7 and 3.8).

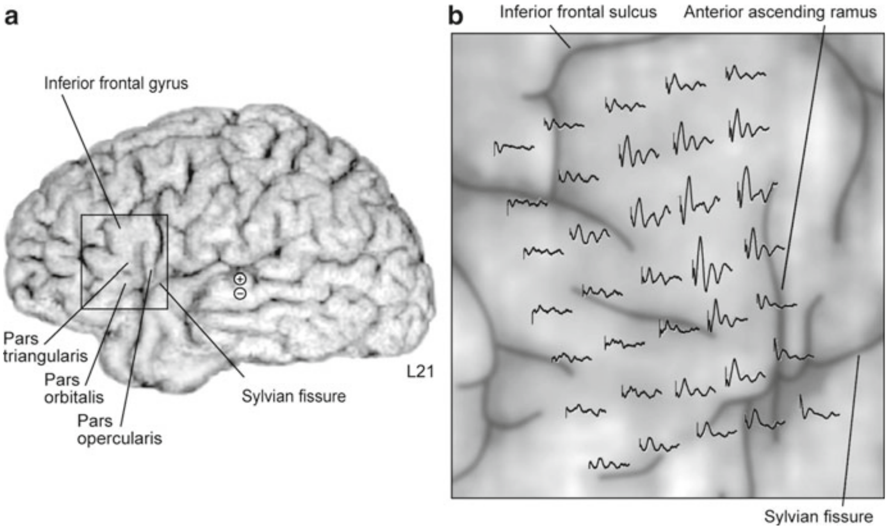


Fig. 3.9 Cortical responses to electrical stimulation of posterolateral STG. (**a**) Lateral view of the left cerebral hemisphere showing stimulation and recording areas (circles and rectangle, respectively). (**b**) Enlarged view of the recording area on prefrontal cortex showing polyphasic electrically-evoked potentials

3.4.2 Functional Connectivity

In nonhuman primates, auditory cortical fields of the temporal lobe and the auditory-related fields of the parietal and frontal lobes are highly interconnected to allow for both serial and parallel processing of acoustic information (Kaas & Hackett, 2005; Rauschecker & Scott, 2009). Anatomical tract tracing methods that have been used so effectively in mapping auditory cortical connectivity in the living monkey brain cannot, however, be used in humans. The use of diffusion tensor imaging has been effective in tracing major white matter tracts in the living human brain (see Talavage, Johnsrude, and Gonzalez Castillo, Chapter 6). One such tract is the arcuate fasciculus that connects temporal cortex with parietal and frontal fields and that, from the time of Wernicke, has been associated with language function and dysfunction. Another includes a pathway coursing through the external capsule and a third reaches orbitofrontal cortex by way of the uncinate fasciculus (Catani et al., 2003, 2005; Glasser & Rilling, 2008). An alternative method of tracing auditory cortical pathways used effectively in the past in animal experiments and more recently in human subjects involves focal electrical stimulation (single charge-balanced 0.1–0.2 ms current pulse) of one cortical site while recording from distant sites (Liegeois-Chauvel et al., 1991; Howard et al., 2000; Brugge et al., 2003; Greenlee et al., 2004; Matsumoto et al., 2004). This method may be applied in acute and chronic situations. In the example illustrated in Figure 3.9, an electrical stimulus applied to auditory cortex on posterolateral STG evoked complex, polyphasic, AEPs that aggregated on ventral prefrontal cortex (vPFC), an area that may be the homolog of vPFC in the macaque monkey that receives a direct anatomical projection from auditory belt and parabelt areas (Hackett et al., 1999; Romanski et al. 1999; Romanski & Goldman-Rakic, 2002). This method is particularly well suited for use in the operating room, as specially designed and fabricated recording and stimulating electrodes may be quickly placed on cortical sites under visual control (see Fig. 3.4), and there is no required action on the part of the subject. Although this approach provides no direct information on the cellular origin or anatomical trajectories of neural pathways, it does give direct information in the living brain on the functional connectivity between the site of electrical stimulation and the site(s) of recording. Using this approach, functional connectivity has been documented between core auditory cortex on HG and an auditory field on posterolateral STG, between that field and the inferior frontal gyrus (IFG), between the IFG and motor cortex of the precentral gyrus, and between subfields within the IFG. Connectivity has also been inferred from patterns of coherence between distant sites as revealed in the electrophysiological recording data (Oya et al., 2007; Gourevitch et al., 2008). Such inferences may be tested recording sound-evoked activity from auditory fields and employing electrical stimulation tract tracing in the same subject.

3.4.3 Electrical Stimulation Functional Mapping

As described earlier, the first experiments carried out in the operating room to study the functional organization of human auditory cortex involved the use of electrical stimulation methods. This approach to create a “functional lesion” by briefly and reversibly disrupting cortical processing in a small cortical area beneath and adjacent to the stimulating electrodes has been further refined. Boatman and her colleagues at Johns Hopkins University have made some of the most effective use of this approach to systematically study auditory cortex functions on the lateral hemispheric surface (Boatman et al., 1995; Boatman, 2004; Sinai et al., 2005; Sinha et al., 2005). These studies are now performed not only in the operating room, but in a more controlled setting associated with chronic recording. Under these conditions, stimulating current (0.3 ms 10–15 mA alternating polarity square wave pulses, 50 Hz, 5–10 s duration) is directed through pairs of adjacent contacts in electrode arrays on the pial surface or in depth electrodes within the supratemporal plane. The approach now can include the use of simultaneous ECoG recordings both to correlate the effects of stimulation with physiological events (Sinai et al., 2005) and to ensure that stimulus intensity does not exceed after-discharge threshold. In chronically implanted subjects controlled psychophysical testing is performed before, and then during periods of electrical stimulation. These experiments have identified sites on lateral STG that appear to be involved in higher order, phonological and lexical-semantic processing of speech (Fig. 3.10), thus providing a framework for a cortical model of speech perception. Electrical stimulation of the STG may also suppress the perception of sound, a phenomenon described originally by Penfield

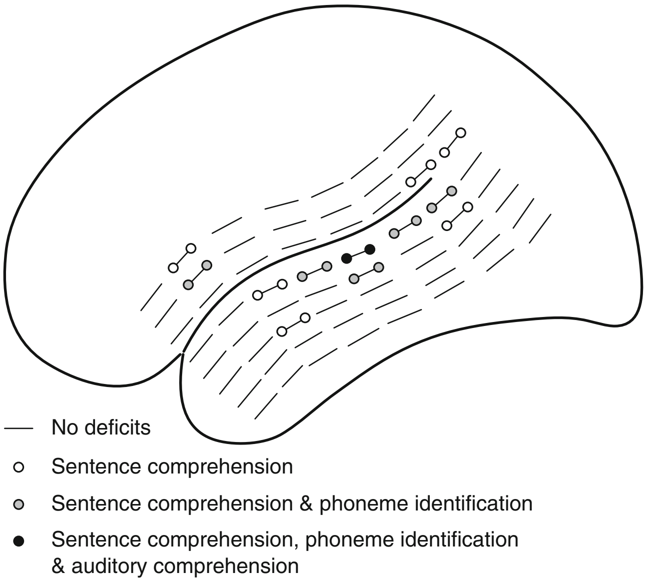

— No deficits
○ Sentence comprehension
● Sentence comprehension & phoneme identification
● Sentence comprehension, phoneme identification & auditory comprehension

Fig. 3.10 Schematic diagram of a cerebral hemisphere showing the locations of auditory discrimination deficits, phoneme identification errors and auditory sentence comprehension deficits induced during electrical stimulation mapping. Electrode locations have been normalized within a standard brain atlas. Lines represent sites where auditory discrimination was tested, but no deficits were induced. (Redrawn from Boatman, 2004.)

and his group as “deafening.” Stimulation of the STG may attenuate ambient sound sensation (Sinha et al., 2005) and, surprisingly, also suppress tinnitus (Fenoy et al., 2006). Another method available to disrupt processing in localized regions of cerebral cortex is local cooling (Bakken et al., 2003). Unlike electrical stimulation, cooling blocks synaptic transmission in the area under the cooling probe. To date, its use is confined to acute experiments performed in the operating room.

3.5 Validity of Invasive Recordings

Although these invasive studies are intended to provide insights into the functional organization of normal human auditory cortex, interpreting the results obtained is done within the context of repeated seizure activity experienced by the study patients, often over periods of years, as well as the present and long-term use of multiple antiepileptic drugs. Surgical patients serving as research subjects are only those diagnosed with a focal seizure disorder. The seizure focus is typically localized to the mesial portion of the temporal lobe, including the hippocampus, where anatomical malformations have been well documented in individuals with drug resistant temporal lobe epilepsy (Gloor, 1991). Thalamic atrophy is reported to be most intense in thalamic nuclei having strong connections with limbic structures

(Bonilha et al., 2006). Epileptic discharges that occasionally invade cortex under study from a distant seizure focus are routinely detected and excluded from analyses. There is also evidence that in addition to mesial temporal malformations there is progressive atrophy of temporopolar, orbitofrontal, insular, and parietal areas (Bernhardt et al., 2009) as well as widespread thinning of neocortex, including lateral temporal regions (Bonilha et al., 2006; Bernhardt et al., 2010) that are considered to be auditory or auditory-related cortex. Thus, one cannot rule out the possibility that pathological processes associated with seizure disorders influence activity recording from distant cortical sites. We also note, however, that data obtained from the auditory core in particular bear a striking resemblance to those recorded from core cortex in the monkey (Fishman et al., 2001; Steinschneider et al., 2005; Ray et al., 2008; Brugge et al., 2009), suggesting that functional organization and certain stimulus–response relationships found in this area have been relatively spared.

3.6 Summary

The methodology of invasive research of human auditory cortex has made tremendous progress since the early studies of Penfield and his colleagues. These developments have paralleled the strides made in developing noninvasive imaging and electrophysiological recording methods. By employing invasive and noninvasive approaches in complementary ways to studies of the functional organization of auditory cortex, the knowledge gained promises to be far greater than that obtainable by relying on any one method alone.

Despite advances, however, technical shortcomings continue to impose limitations on invasive cortical electrophysiological recording and stimulation as research tools. Chronically implanted electrode arrays are tethered to head-mounted connectors that are, in turn, connected to external electronic instruments. Current data acquisition systems and surgical techniques allow for extensive (>200 contacts) electrode coverage making the external cables bulky and sometimes uncomfortable for the patient. External electrical (power line) noise easily coupled to wires often introduces unwanted interference during recording sessions. Wires can, and do, break, especially during seizures, resulting in loss of both clinical and research data. All of these considerations have a direct impact on the conduct and outcome of research, which in turn relate directly to patient safety and successful diagnosis of brain disorders and to successful development of neural prostheses.

Solutions will come through advances in engineering and material science. Miniaturization and tailoring of implanted arrays will be found in thin-film technology, currently in use for electronic circuit design and fabrication, as well as emerging nanotechnology coupled, perhaps, with the aid of magnetic navigation for implanting miniaturized and flexible depth electrodes. Bioactive conductive polymers may replace metal as material for electrode contacts, thereby eliminating concerns over potential electrochemical tissue damage. Finally, replacing bulky cables with wireless transmission will allow clinical and research data to be obtained under a far wider range of environmental conditions. Taken together, these, and other, future technical advances will enhance patient comfort and safety, improve diagnosis and treatment, and open new opportunities for research.

Acknowledgments Preparation of this chapter was supported by NIH RO1-DC004290, UL1RR024979, and GCRC MO1-RR-59 and by the Hoover Fund.

References

Bakken, H. E., Kawasaki, H., Oya, H., Greenlee, J. D., & Howard, M. A. (2003). A device for cooling localized regions of human cerebral cortex. *Journal of Neurosurgery*, 99, 604–608.

Bernhardt, B. C., Worsley, K. J., Kim, H., Evans, A. C., Bernasconi, A., & Bernasconi, N. (2009). Longitudinal and cross-sectional analysis of atrophy in pharmacoresistant temporal lobe epilepsy. *Neurology*, 72, 1747–1754.

Bernhardt, B. C., Bernasconi, N., Concha, L., & Bernasconi, A. (2010). Cortical thickness analysis in temporal lobe epilepsy: Reproducibility and relation to outcome. *Neurology*, 74(22), 1776–1784.

Bitterman, Y., Mukamel, R., Malach, R., Fried, I., & Nelken, I. (2008). Ultra-fine frequency tuning revealed in single neurons of human auditory cortex. *Nature*, 451(7175), 197–201.

Boatman, D. (2004). Cortical bases of speech perception: Evidence from functional lesion studies. *Cognition*, 92, 47–65.

Boatman, D., Lesser, R. P., & Gordon, B. (1995). Auditory speech processing in the left temporal lobe: An electrical interference study. *Brain and Language*, 51(2), 269–290.

Bonilha, L., Rorden, C., Appenzeller, S., Coan, A. C., Cendes, F., & Li, L. M. (2006). Gray matter atrophy associated with duration of temporal lobe epilepsy. *NeuroImage*, 32, 1070–1079.

Brugge, J. F., Volkov, I. O., Garell, P. C., Reale, R. A., & Howard, M. A. (2003). Functional connections between auditory cortex on Heschl's gyrus and on the lateral superior temporal gyrus in humans. *Journal of Neurophysiology*, 90, 3750–3763.

Brugge, J. F., Volkov, I. O., Reale, R. A., Garell, P. C., Kawasaki, H., Oya, H., et al. (2005). The posteriolateral superior temporal auditory field in humans. Functional organization and connectivity. In R. Konig, P. Heil, E. Budinger, & H. Scheich (Eds.), *The auditory cortex—toward a synthesis of human and animal research* (pp. 145–162). Mahwah, NJ: Erlbaum.

Brugge, J. F., Volkov, I. O., Oya, H., Kawasaki, H., Reale, R. A., Fenoy, A., et al. (2008). Functional localization of auditory cortical fields of human: Click-train stimulation. *Hearing Research*, 238(1–2), 12–24.

Brugge, J. F., Nourski, K. V., Oya, H., Reale, R. A., Kawasaki, H., Steinschneider, M., & Howard, M. A. (2009). Coding of repetitive transients by auditory cortex on Heschl's gyrus. *Journal of Neurophysiology*, 102(4), 2358–2374.

Catani, M., Jones, D. K., Donato, R., & Ffytche, D. H. (2003). Occipito-temporal connections in the human brain. *Brain*, 126(Pt 9), 2093–2107.

Catani, M., Jones, D. K., & ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Annals of Neurology*, 57(1), 8–16.

Celesia, G. G. (1976). Organization of auditory cortical areas in man. *Brain*, 99, 403–414.

Celesia, G. G., & Puletti, F. (1969). Auditory cortical areas of man. *Neurology*, 19, 211–220.

Cohen, Y. E., Russ, B. E., Davis, S. J., Baker, A. E., Ackelson, A. L., & Nitecki, R. (2009). A functional role for the ventrolateral prefrontal cortex in non-spatial auditory cognition. *Proceedings of the National Academy of Sciences of the USA*, 106(47), 20045–20050.

Creutzfeldt, O., & Ojemann, G. (1989). Neuronal activity in the human lateral temporal lobe. III. Activity changes during music. *Experimental Brain Research*, 77(3), 490–498.

Creutzfeldt, O., Ojemann, G., & Lettich, E. (1989a). Neuronal activity in the human lateral temporal lobe. I. Responses to speech. *Experimental Brain Research*, 77(3), 451–475.

Creutzfeldt, O., Ojemann, G., & Lettich, E. (1989b). Neuronal activity in the human lateral temporal lobe. II. Responses to the subjects own voice. *Experimental Brain Research*, 77(3), 476–489.

Crone, N. E., Boatman, D., Gordon, B., & Hao, L. (2001). Induced electrocorticographic gamma activity during auditory perception. *Clinical Neurophysiology*, 112(4), 565–582.

Crone, N. E., Sinai, A., & Korzeniewska, A. (2006). High-frequency gamma oscillations and human brain mapping with electrocorticography. *Progress in Brain Research*, 159, 275–295.

Davis, P. A. (1939). Effects of acoustic stimuli on the waking human brain. *Journal of Neurophysiology*, 2, 494–499.

Edwards, E., Soltani, M., Kim, W., Dalal, S. S., Nagarajan, S. S., Berger, M. S., & Knight, R. T. (2009). Comparison of time-frequency responses and the event-related potential to auditory speech stimuli in human cortex. *Journal of Neurophysiology*, 102(1), 377–386.

Engel, A. K., König, P., Kreiter, A. K., Singer, W. (1991). Interhemispheric synchronization of oscillatory neuronal responses in cat visual cortex. *Science*, *252*(5010), 1177–1179.

Engel, A. K., Moll, C. K., Fried, I., & Ojemann, G. A. (2005). Invasive recordings from the human brain: Clinical insights and beyond. *Nature Reviews Neuroscience*, 6(1), 35–47.

Engel, J. J. (2001). Finally, a randomized, controlled trial of epilepsy surgery. *New England Journal of Medicine*, 345, 365–367.

Fenoy, A. J., Severson, M. A., Volkov, I. O., Brugge, J. F., & Howard, M.A. (2006). Hearing suppression induced by electrical stimulation of human auditory cortex. *Brain Research*, 1118, 75–83.

Fishman, Y. I., Volkov, I. O., Noh, M. D., Garell, P. C., Bakken, H., Arezzo, J. C., et al. (2001). Consonance and dissonance of musical chords: Neural correlates in auditory cortex of monkeys and humans. *Journal of Neurophysiology*, 86(6), 2761–2788.

Flinker, A., Chang, E. F., Barbaro, N. M., Berger, M. S., & Knight, R. T. (2011). Sub-centimeter language organization in the human temporal lobe. *Brain and Language*. doi: S0093–934X(10)00155–0 [pii].

Freeman, W. J. (1978). Spatial properties of an EEG event in the olfactory bulb and cortex. *Electroencephalography and Clinical Neurophysiology*, 44, 586–605.

Fried, I., Wilson, C. L., Maidment, N. T., Engel, J., Jr., Behnke, E., Fields, T. A., et al. (1999). Cerebral microdialysis combined with single-neuron and electroencephalographic recording in neurosurgical patients. Technical note. *Journal of Neurosurgery*, 91(4), 697–705.

Geisler, C. D., Frishkopf, L. S., & Rosenblith, W. A. (1958). Extracranial responses to acoustic clicks in man. *Science*, 128, 1210–1211.

Ghovanloo, M., Otto, K. J., Kipke, D. R., & Najafi, K. (2004). In vitro and in vivo testing of a wireless multichannel stimulating telemetry microsystem. *Proceedings of the 26th Annual International Conference of the IEEE EMBS*, 6, 4294–4297.

Glasser, M. F., & Rilling, J. K. (2008). DTI tractography of the human brain's language pathways. *Cerebral Cortex*, 18(11), 2471–2482.

Gloor, P. (1991). Mesial temporal sclerosis: Historical background and an overview from a modern perspective. In H. O. Luders (Ed.), *Epilepsy surgery* (pp. 689–703). New York: Raven Press.

Gloor, P., Olivier, A., Quesney, L. F., Andermann, F., & Horowitz, S. (1982). The role of the limbic system in experiential phenomena of temporal lobe epilepsy. *Annals of Neurology*, 12, 129–144.

Gourevitch, B., Le Bouquin Jeannes, R., Faucon, G., & Liegeois-Chauvel, C. (2008). Temporal envelope processing in the human auditory cortex: Response and interconnections of auditory cortical areas. *Hearing Research*, 237(1–2), 1–18.

Gray, C. M., König, P., Engel, A. K., & Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, 338, 334–337.

Greenlee, J. D., Oya, H., Kawasaki, H., Volkov, I. O., Kaufman, O. P., Kovach, C., et al. (2004). A functional connection between inferior frontal gyrus and orofacial motor cortex in human. *Journal of Neurophysiology*, 92(2), 1153–1164.

Greenlee, J. D., Oya, H., Kawasaki, H., Volkov, I. O., Severson, M. A., 3rd, Howard, M. A., 3rd, & Brugge, J. F. (2007). Functional connections within the human inferior frontal gyrus. *Journal of Comparative Neurology*, 503(4), 550–559.

Griffiths, T. D., Kumar, S., Sedley, W., Nourski, K. V., Kawasaki, H., Oya, H., et al. (2010). Direct recordings of pitch responses from human auditory cortex. *Current Biology*, 20(12), 1128–1132.

Hackett, T. A. (2003). The comparative anatomy of the primate auditory cortex. In A. A. Ghazanfar (Ed.), *Primate audition: Ethology and neurobiology* (pp. 199–219). Boca Raton: CRC Press.

Hackett, T. A. (2007). Organization and correspondence of the auditory cortex of humans and nonhuman primates. In J. H. Kaas (Ed.), *Evolution of the nervous system* (pp. 109–119). Oxford: Elsevier.

Hackett, T. A. (2008). Anatomical organization of the auditory cortex. *Journal of the American Academy of Audiology*, 19(10), 774–779.

Hackett, T. A., Stepniewska, I., & Kaas, J. H. (1999). Prefrontal connections of the parabelt auditory cortex in macaque monkeys. *Brain Research*, 817(1–2), 45–58.

Howard, M. A., Volkov, I. O., Abbas, P. J., Damasio, H., Ollendieck, M. C., & Granner, M. A. (1996a). A chronic microelectrode investigation of the tonotopic organization of human auditory cortex. *Brain Research*, 724, 260–264.

Howard, M. A., Volkov, I. O., Granner, M. A., Damasio, H. M., Ollendieck, M. C., & Bakken, H. E. (1996b). A hybrid clinical-research depth electrode for acute and chronic in vivo microelectrode recording of human brain neurons. Technical note. *Journal of Neurosurgery*, 84, 129–132.

Howard, M. A., Volkov, I. O., Mirsky, R., Garell, P. C., Noh, M. D., Granner, M., et al. (2000). Auditory cortex on the posterior superior temporal gyrus of human cerebral cortex. *Journal of Comparative Neurology*, 416, 76–92.

Jenison, R. L., Rangel, A., Oya, H., Kawasaki, H., & Howard, M. A. (2011). Value encoding in single neurons in the human amygdala during decision making. *Journal of Neuroscience*, 31(1), 331–338.

Johnson, M. D., Otto, K. J., Williams, J. C., & Kipke, D. R. (2004). Bias voltages at microelectrodes change neural interface properties in vivo. *Proceedings of the 26th Annual International Conference of the IEEE EMBS*, 6, 4103–4106.

Kaas, J. H., & Hackett, T. A. (2005). Subdivisions and connections of auditory cortex in primates: A working model. In R. Konig, P. Heil, E. Budinger, & H. Scheich (Eds.), *Auditory cortex. A synthesis of human and animal research* (pp. 7–25). Mahwah, NJ: Erlbaum.

Kellis, S., Miller, K., Thomson, K., Brown, R., House, P., & Greger, B. (2010). Decoding spoken words using local field potentials recorded from the cortical surface. *Journal of Neural Engineering*, 7(5), 056007.

Lempka, S. F., Johnson, M. D., Barnett, D. W., Moffitt, M. A., Otto, K. J., Kipke, D. R., & McIntyre, C. C. (2006). Optimization of microelectrode design for cortical recording based on thermal noise considerations. *Proceedings of the 28th IEEE EMBS Annual International Conference*, 1, 3361–3364.

Leonard, C. M., Puranik, C., Kuldau, J. M., & Lombardino, L. J. (1998). Normal variation in the frequency and location of human auditory cortex landmarks. Heschl's gyrus: Where is it? *Cerebral Cortex*, 8, 397–406.

Liégeois-Chauvel, C., Musolino, A., & Chauvel, P. (1991). Localization of the primary auditory area in man. *Brain*, 114, 139–151.

Liégeois-Chauvel, C., Musolino, A., Badier, J. M., Marquis, P., & Chauvel, P. (1994). Evoked potentials recorded from the auditory cortex in man: Evaluation and topography of the middle latency components. *Electroencephalography and Clinical Neurophysiology*, 92, 204–214.

Liégeois-Chauvel, C., de Graaf, J. B., Laguitton, V., & Chauvel, P. (1999). Specialization of left auditory cortex for speech perception in man depends on temporal coding. *Cerebral Cortex*, 9, 484–496.

Liégeois-Chauvel, C., Lorenzi, C., Trebuchon, A., Regis, J., & Chauvel, P. (2004). Temporal envelope processing in the human left and right auditory cortices. *Cerebral Cortex*, 14(7), 731–740.

Lu, T., Liang, L., & Wang, X. (2001). Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. *Nature Neuroscience*, 4(11), 1131–1138.

Matsumoto, R., Nair, D. R., LaPresto, E., Najm, I., Bingaman, W., Shibasaki, H., & Luders, H. O. (2004). Functional connectivity in the human language system: A cortico-cortical evoked potential study. *Brain*, 127(Pt 10), 2316–2330.

Matsumoto, R., Nair, D. R., LaPresto, E., Bingaman, W., Shibasaki, H., & Luders, H. O. (2007). Functional connectivity in human cortical motor system: A cortico-cortical evoked potential study. *Brain*, 130(Pt 1), 181–197.

Miller, K. J., Makeig, S., Hebb, A. O., Rao, R. P., denNijs, M., & Ojemann, J. G. (2007). Cortical electrode localization from X-rays and simple mapping for electrocorticographic research: The "Location on Cortex" (LOC) package for MATLAB. *Journal of Neuroscience Methods*, 162(1–2), 303–308.

Moriarity, J. L., Boatman, D., Krauss, G. L., Storm, P. B., & Lenz, F. A. (2001). Human "memories" can be evoked by stimulation of the lateral temporal cortex after ipsilateral medial temporal lobe resection. *Journal of Neurology*, *Neurosurgery and Psychiatry*, 71(4), 549–551.

Mukamel, R., Gelbard, H., Arieli, A., Hasson, U., Fried, I., & Malach, R. (2005). Coupling between neuronal firing, field potentials, and FMRI in human auditory cortex. *Science*, 309(5736), 951–954.

Nourski, K. V., Reale, R. A., Oya, H., Kawasaki, H., Kovach, C. K., Chen, H., et al. (2009). Temporal envelope of time-compressed speech represented in the human auditory cortex. *Journal of Neuroscience*, 29(39), 15564–15574.

Oya, H., Poon, P. W., Brugge, J. F., Reale, R. A., Kawasaki, H., Volkov, I. O., & Howard, M. A., 3rd. (2007). Functional connections between auditory cortical fields in humans revealed by Granger causality analysis of intra-cranial evoked potentials to sounds: comparison of two methods. *Biosystems*, 89(1–3), 198–207.

Penfield, W., & Perot, P. (1963). The brain's record of auditory and visual experience—a final summary and discussion. *Brain*, 86, 595–696.

Penfield, W., & Rasmussen, T. (1950). *The cerebral cortex of man—A clinical study of localization of function*. New York: Macmillan.

Pierce, A. L., Sommakia, S., Rickus, J. L., & Otto, K. J. (2009). Thin-film silica sol-gel coatings for neural microelectrodes. *Journal of Neuroscience Methods*, 180(1), 106–110.

Preuss, T. M. (1995). The argument from animals to humans in cognitive neuroscience. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 1227–1241). Cambridge, MA: MIT Press.

Rademacher, J., Caviness, V., Steinmetz, H., & Galaburda, A. (1993). Topographical variation of the human primary cortices; implications for neuroimaging, brain mapping and neurobiology. *Cerebral Cortex*, 3, 313–329.

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6), 718–724.

Ray, S., Niebur, E., Hsiao, S. S., Sinai, A., & Crone, N. E. (2008). High-frequency gamma activity (80–150 Hz) is increased in human cortex during selective attention. *Clinical Neurophysiology*, 119(1), 116–133.

Reddy, C. G., Dahdaleh, N.S., Albert, G., Chen, F., Hansen, D., Nourski, K., et al. (2010). A method for placing Heschl gyrus depth electrodes. *Journal of Neurosurgery*, 112(6), 1301–1307.

Romanski, L. M. (2004). Domain specificity in the primate prefrontal cortex. Cognitive, Affective, & Behavioral Neuroscience, 4(4), 421–429.

Romanski, L. M., & Goldman-Rakic, P. S. (2002). An auditory domain in primate prefrontal cortex. *Nature Neuroscience*, 5(1), 15–16.

Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., & Rauschecker, J. P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nature Neuroscience*, 2, 1131–1136.

Rosen, S. (1992). Temporal information in speech: Acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 336(1278), 367–373.

Sinai, A., Bowers, C. W., Crainiceanu, C. M., Boatman, D., Gordon, B., Lesser, R. P., et al.(2005). Electrocorticographic high gamma activity versus electrical cortical stimulation mapping of naming. *Brain*, 128(Pt 7), 1556–1570.

Sinha, S. R., Crone, N. E., Fotta, R., Lenz, F., & Boatman, D. F. (2005). Transient unilateral hearing loss induced by electrocortical stimulation. *Neurology*, 64, 383–385.

Steinschneider, M., Reser, D. H., Fishman, Y. I., Schroeder, C. E., & Arezzo, J. C. (1998). Click train encoding in primary auditory cortex of the awake monkey: Evidence for two mechanisms subserving pitch perception. *Journal of the Acoustical Society of America*, 104(5), 2935–2955.

Steinschneider, M., Volkov, I. O., Fishman, Y. I., Oya, H., Arezzo, J. C., & Howard, M. A., 3rd. (2005). Intracortical responses in human and monkey primary auditory cortex support a temporal processing mechanism for encoding of the voice onset time phonetic parameter. *Cerebral Cortex*, 15(2), 170–186.

Steinschneider, M., Fishman, Y. I., & Arezzo, J. C. (2008). Spectrotemporal analysis of evoked and induced electroencephalographic responses in primary auditory cortex (A1) of the awake monkey. *Cerebral Cortex*, 18(3), 610–625.

World Health Organization (2005). *Atlas: Epilepsy care in the world*. Geneva: WHO.

Chapter 4
Recording Event-Related Brain Potentials: Application to Study Auditory Perception

Claude Alain and István Winkler

4.1 Introduction

The last decade has seen an explosion of research in auditory perception and cognition. This growing activity encompasses neurophysiological research in nonhuman species, computational modeling of basic neurophysiological functions, and neuroimaging research in humans. Among the various neuroimaging techniques available, scalp recording of neuroelectric (electroencephalography [EEG]) and neuromagnetic (magnetoencephalography [MEG]) (see Nagarajan, Gabriel, and Herman, Chapter 5) brain activity have proven to be formidable tools in the arsenal available to cognitive neuroscientists interested in understanding audition. These techniques measure the dynamic pattern of electromagnetic fields at the scalp produced by the coherent activity of large neuronal populations in the brain. In cognitive neuroscience, the measurement of the electrical event-related brain potentials (ERPs) or magnetic event-related fields (ERFs) is among the major noninvasive techniques used for investigating sensory and cognitive information processing and for testing specific assumptions of cognitive theories that are not easily amenable to behavioral techniques. After identifying and characterizing the ERP/ERF signals that accompany the basic steps of processing discrete events, scientific interest has

C. Alain (✉)
Rotman Research Institute, Baycrest Centre for Geriatric Care, 3560 Bathurst Street,
Toronto, Ontario M6A 2E1, Canada

Department of Psychology, University of Toronto, Ontario M8V 2S4, Canada
e-mail: calain@rotman-baycrest.on.ca

I. Winkler
Institute for Psychology, Hungarian Academy of Sciences,
P.O. Box 398, H-1394 Budapest, Hungary

Institute of Psychology, University of Szeged, 6722 Szeged, Petőfi S. sgt. 30-34, Hungary
e-mail: iwinkler@cogpsyphy.hu

D. Poeppel et al. (eds.), *The Human Auditory Cortex*, Springer Handbook
of Auditory Research 43, DOI 10.1007/978-1-4614-2314-0_4,

gradually shifted toward specifying the complex processing of more realistic stimulus configurations. In the auditory modality, recent years have seen an upsurge of research papers investigating the processes of auditory scene analysis (ASA) by ERP/ERF methods (for recent reviews, see Alain, 2007; Snyder & Alain, 2007; Winkler et al., 2009a).

This chapter discusses the contributions of ERPs (and ERFs) to our understanding of auditory cognition, in general, and ASA, in particular. Some concepts relevant to the generation, recording, and analysys of EEG data are first reviewed. Work from Luck (2005) and Picton (2010) can be consulted for more extensive reviews of EEG recording and analysis. The main concepts relevant to ASA are reviewed next, followed by examples illustrating the usefulness of ERPs in auditory cognitive neuroscience. We conjecture that although auditory predictions are formed and the auditory scene is resolved to a certain degree outside the focus of attention, attention modulates several processes of ASA and usually determines the object representations appearing in conscious perception. We conclude by proposing future research directions.

4.2 Recording of Neuroelectric Brain Activity

Scalp recordings of EEG and ERPs have endured the test of time and have proven to be an important tool in investigating the psychological and neural processes underlying the perception and formation of auditory objects in humans. It all began in 1924, when Hans Berger demonstrated that it was possible to record small electrical responses from the brain using sensors (electrodes) attached to the scalp. Although the scientific community was at first skeptical, by 1934 the notion that neuroelectric brain activity could be recorded with electrodes placed on the scalp had been established. The subsequent years saw the development and validation of this "new" imaging technique as a powerful clinical tool in neurology. While EEG was gaining momentum in clinical practice, evidence was also mounting revealing changes in the ongoing EEG synchronized to the onset of visual or auditory stimuli. Hans Berger was likely the first to notice changes in alpha rhythm in response to a sudden sound, though to our knowledge the research on auditory ERPs truly began with the paper from Davis (1939), who showed time-locked changes in the ongoing EEG to various sounds. It was not until the widespread availability of the digital computer that auditory ERPs became widely used in research.

The neuroelectric signals recorded at the scalp reflect the summation of excitatory and inhibitory postsynaptic potentials from large numbers of pyramidal neurons whose parallel geometric configuration allows for effective summation of postsynaptic potentials. These postsynaptic potentials are very small (millionth of a volt [μV]) and typically need to be amplified at the order of 10,000 to 1,000,000 to be observable at the scalp. The EEG records such "brain waves" by computing the voltage difference between two electrodes. The spatial configuration of the electrodes used for recording voltage differences is referred to as a "montage." In a bipolar montage, typically used in clinical settings (e.g., testing for epilepsy), the EEG is obtained by comparing two adjacent electrodes that are often considered “active” because they both “capture” brain activity. In a monopolar montage (also referred to referential montage), which is more popular in research, one electrode is considered “active” while the other is thought to be “passive” or “inactive.” In reality, a passive or inactive electrode does not exist, and one needs to choose the reference electrode appropriately (i.e., an electrode that is not involved in the electrical field being studied). Commonly used reference electrodes in auditory research are located at the mastoids (often linked together), the nose, or a balanced noncephalic reference (e.g., the shoulder or neck). More recently, the common average reference (subtracting the average activity across all electrode locations) has become a popular choice. However, it is appropriate only when a large number of electrodes are used and when they cover the whole scalp evenly, including the lower areas near the eyes, ear, and the cerebellar region.

The electrical activity recorded at the scalp is often described in terms of equivalent current dipoles. These representations closely approximate the parallel organization of the pyramidal neurons. The assumption is that the synchronized activity of these pyramidal neurons can be modeled by a point-like electrical source (dipole). Radial dipoles are perpendicular to the skull and likely formed by activation of postsynaptic potentials in the gyral cortex whereas tangential dipoles are parallel to the skull and reflect postsynaptic potentials in sulci. The activity in auditory cortex is best modeled by both tangential and radial sources along the superior temporal plane near Heschl’s gyrus (Picton et al., 1999).

There are several methodological details that one needs to consider when planning a study that involves EEG recording. These include the sampling rate of the EEG data, filter settings, the reference electrode, the spatial configuration of electrodes, and the amplification and filtering of the ongoing EEG signal. The choice of the sampling rate is determined by the type of response of interest and it should be at least twice the frequency of the signal being studied (i.e., the Nyquist rate). For instance, the sampling rate used in most studies of auditory perception and memory typically varies between 200 and 500 Hz because most researchers are interested in long latency evoked potentials (which have slow time constants and do not require a high sampling rate). However, there is increasing evidence that complex sounds may be represented in lower auditory centers. Support for this comes from studies that have used higher sampling rates (e.g., 1–10 kHz) to examine brain stem evoked responses and frequency-following responses (Picton, 2010). The latter has a waveform similar to that of the stimulus and is thought to reflect low-level representations in the ascending auditory pathway. When in doubt, one should consider sampling the EEG data at a higher rate because one can always down sample (i.e., decimate) the data, whereas a lack of sufficient temporal resolution may mask or distort the results. During recording, the filter should be set at half the sampling rate to ensure that no signal exceeds the Nyquist rate. Signals with a very low rate of change (slow signals, termed “direct current” potentials) also carry important information about psychologically relevant processes in the brain. For example, preparatory potentials may extend to several seconds, resulting in signal frequencies of 0.1–1 Hz. Unfortunately, such slow signals can cause recording problems because their amplitude is often much higher than that of the faster ones. Unless an appropriate

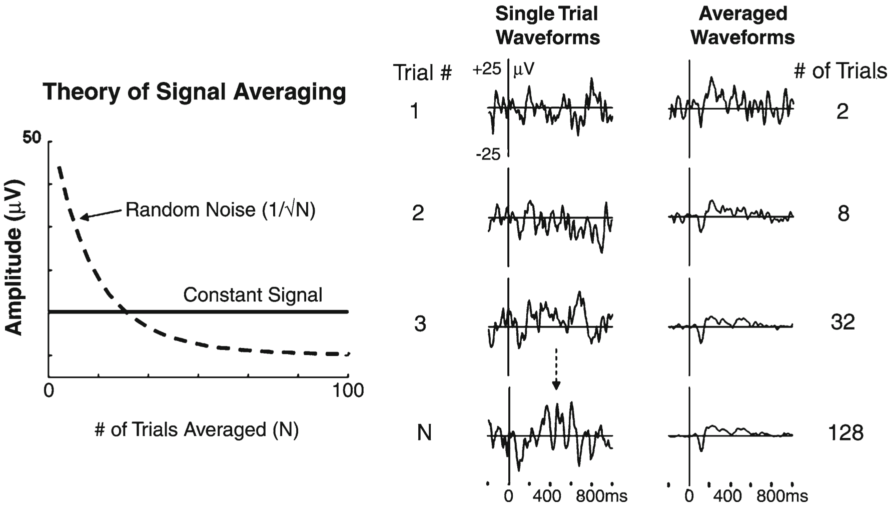


Fig. 4.1 Averaging. The left panel shows the theory of signal averaging intended to increase the strength of a constant signal relative to noise. The right panel shows the single trial data and the changes in signal-to-noise ratio as a function of the number of trials used to form the average. The data come from an auditory oddball paradigm and show the response to the frequent (standard) stimulus

amplifier is used and electrodes are carefully mounted on the scalp, it is advisable to filter out the slow signals, for example, by setting a high-pass filter with a 0.01 Hz limit. As a consequence, in most studies of long latency evoked potentials, a band-pass filter is set between 0.01 and 50 or 100 Hz.

At present, in most laboratories, EEG is recorded using multichannel recording systems that comprise 16, 32, 64, 128, 256 or even up to 512 electrodes. The number of electrodes needed to define the scalp distribution of an ERP depends on how rapidly the potential changes over the scalp. The rule of thumb is to set up at least two electrodes within one cycle of the highest spatial frequency within the scalp-recorded fields. An initial high-density recording can indicate the highest spatial frequency in the ERP waveform of interest and this can then be used to determine the minimum number of electrodes for subsequent recordings. The electrode locations are based on four anatomical landmarks: the nasion (indentation between the forehead and the nose), the inion (ridge that can be felt at the midline of the back of the skull), and the left and right preauricular points (indentations near the external ear openings). The 10–10 or 10–20 system refers to the distances (in percentage) between adjacent electrodes.

The most common data analysis technique for separating ERPs from the background neuroelectric activity consists of averaging epochs of brain waves that are time locked to an external sensory event (e.g., tone, speech sound). The basic assumptions are that (1) the signal and noise linearly sum together to produce the recorded waveform, (2) the transient event-related response (i.e., signal waveform) is relatively stable in time from trial to trial, and (3) the "noise" (i.e., unrelated neuroelectric activity) fluctuates randomly from trial to trial such that it can be considered as random samples of a stable (stationary) stochastic process (Fig. 4.1). Although these principles hold in many cases, there are also rapid physiological changes that reflect habituation and learning (Alain et al., 2007, 2010) violating the second assumption. These effects may be more important for long latency responses, which are often modulated by perceptual and attentional states. Other problems may arise from long sessions, during which the participant's arousal state may change (violating assumption 3). As a rule of thumb, averaging across data collected from periods exceeding an hour may be suspect for this violation. Averaging single-trial responses is also quite sensitive to high-amplitude transient electrical signals, such as eye movements (violating assumption 3). This problem can be reduced by (a) excluding trials with high-amplitude transients from the data set (termed artifact rejection), or (b) modeling the sources of such artifacts and separating them from the signal before averaging (e.g., eye-movement correction methods), or (c) calculating the median instead of the average, as the former is less sensitive to high-amplitude artifacts. Finally, the relation between the continuous rhythmic background EEG activity and ERPs is largely unknown. Current theories range from assuming a large degree of independence between the two to claims that transient (ERP) activity emerges as a result of phase synchronization of the continuous rhythmic activity (Picton, 2010). Thus, the extent to which assumption 1 can be relied upon in a given experiment is not known. When in doubt, one should analyze the time course of phase in various spectral bands of the EEG signal with respect to the timing of the auditory stimuli. Despite all these problems, the averaging technique has been, and is still, the most widely used method for extracting ERPs and most of our knowledge about ERPs has been obtained from studies employing it.

The averaged auditory ERPs typically consist of peaks and valleys that reflect synchronous activity from large neuronal ensembles at particular latencies following sound onset. The ERPs measured over the scalp are the linear sum of neuroelectric signals produced by multiple intracerebral sources that overlap in time. Some of these brain waves are exogenous (i.e., obligatory and stimulus driven), in the sense that they occur regardless of the observer's intention and motivation, predominantly reflecting the physical characteristics of the external events. Other brain waves are endogenous because they are co-determined by stimulus properties (e.g., pitch, intensity, duration) and psychological factors such as attention and expectation. The presence and characteristics of ERPs in terms of latency and amplitude are used to make inferences regarding the underlying psychological processes as well as the likely site of origin. Recent advances in brain electrical and magnetic source analysis (see later) have improved our understanding of the neural generators of both sensory (exogenous) and cognitive (endogenous) evoked responses to auditory stimuli, making this technique ideal for assessing the impact of attention and learning on ASA, in general, and object formation, in particular.

The latency of the above described brain waves refers to the amount of time (typically, in milliseconds) that is taken to generate the bioelectrical response following the onset of the event. The resolution of these latency measures is directly related to the sampling rate of the signal. Brain wave amplitude is affected by the strength of the response, the size of the neural population, how well the neuronal activity is synchronized, and where these neurons are located in the brain with respect to the point of measurement. Transient auditory events elicit early brain stem responses (1–10 ms

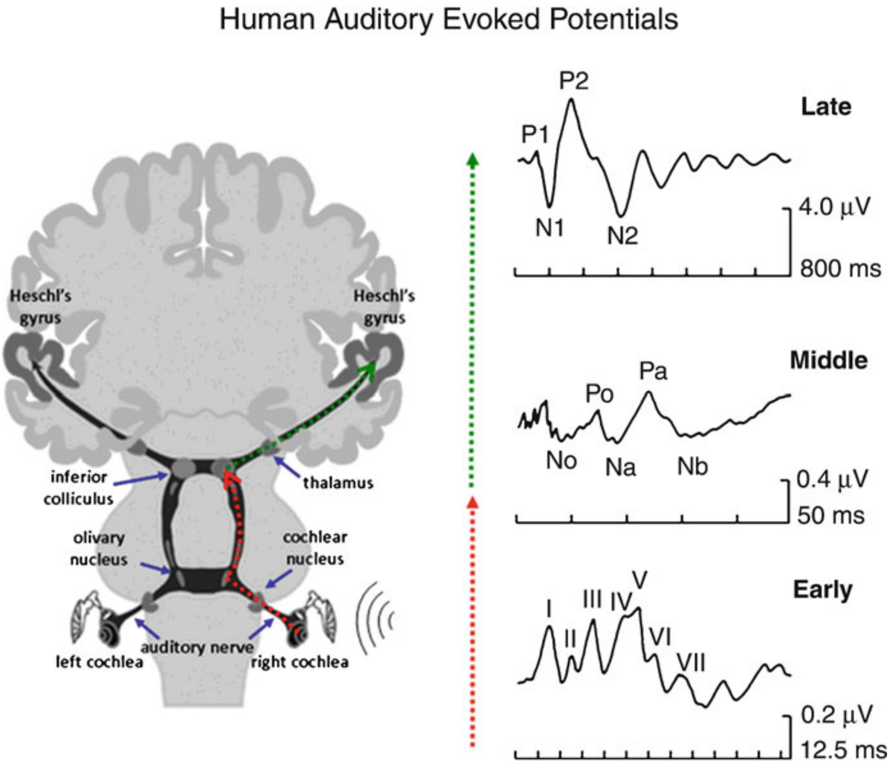


Fig. 4.2 The left side of the figure shows a schematic representation of the ascending auditory pathways. The right side shows the time course of auditory event-related potentials recorded over the midline central scalp region (i.e., Cz) with the reference at the right mastoid. The averaged evoked potentials comprise 1536 responses elicited by clicks presented one per second to the right ear with an intensity of 60 dB hearing level (i.e., 60 dB above pure tone sensation thresholds). The brain stem responses (wave I, II, III, IV, V, VI, and VII), middle latency evoked responses (No, Po, Na, Pa, and Nb), and long latency evoked responses (P1, N1, P2, and N2) are shown in the bottom, middle, and top panel, respectively. The colored arrow shows the progression from brain stem to long latency evoked potentials along the ascending auditory pathways. (Adapted from Picton, 2010.)

after sound onset) followed by the first volley of cortical activity in primary auditory cortex (middle-latency evoked responses, 10–50 ms) and on to later ("long-latency," >50 ms) waves that reflect higher-order auditory processes associated with perception, memory, and other cognitive processes (Fig. 4.2). The long-latency auditory evoked potentials begin with the P1, N1, and P2 waves (Näätänen & Picton, 1987). The letters P and N refer to the polarity of the wave, positive and negative, respectively, at the typical measurement location, which is, in most cases, over the central scalp. Though the polarity of ERP waveforms is a defining feature, it often bears little functional or neurophysiological significance because the polarity of ERP deflections can vary with the location of the reference electrode, the baseline against which it is compared, and the location and orientation of its intracerebral generators. There are three conventions with respect to naming the ERP waves: (a) serial order numbers within the long-latency responses (e.g., P3; further letters refer to subtypes, such as P3b), or (b) the typical or actual peak latency of the response (e.g., P3 is also often referred to as P300, because of its typical peak latency), or (c) the abbreviation of the functional characterization of the response (e.g., the object-related negativity [ORN]). Magnetic counterparts of the electrical ERP responses (i.e., transient magnetic activity accompanying the electric potential changes according to the laws of electromagnetic fields; see Nagarjan, Gabriel, and Herman, Chapter 5) are conventionally marked by the letter "m" appended to the label of the corresponding ERP (e.g., the magnetic counterpart of the N1 ERP is often termed "N1m").

P1, N1, and P2 are particularly sensitive to the signal-to-noise ratio, and usually decrease in amplitude when the signal is embedded in white noise or multitalker babble, though in some circumstances soft Gaussian noise presented in the background can enhance the N1m amplitude (Alain et al., 2009a). The conscious detection of a task-relevant sound (i.e., target) is usually reflected by a negative wave peaking between 150 and 250 ms, termed the N2b (Ritter & Ruchkin, 1992). The N2b is followed by a late positive wave peaking between 250 and 600 ms poststimulus, the P3b, which may reflect working memory and context updating (Verleger, 1988; Picton, 1992; Polich, 2007). In addition to these brain waves, two other brain responses are particularly relevant to the following discussion of auditory cognition and ASA, namely, the mismatch negativity (MMN) and the ORN. Whereas the former is elicited by infrequent violations of the regular features of an ongoing stream of sounds (Näätänen et al., 1978; Picton et al., 2000; Kujala et al., 2007), the latter is generated in situations where more than one sound is heard simultaneously (Alain et al., 2001; Alain, 2007). The relevance of these two brain responses, as they pertain to auditory cognition, is discussed in subsequent sections.

The problem of determining the intracerebral sources for neuroelectric activities measured outside the head (i.e., source localization of electromagnetic brain activity) is referred to as the bioelectromagnetic inverse problem. The inverse problem describes the fact that electromagnetic brain activity measured outside the head does not contain sufficient information for inferring from it a unique source configuration. There are several methods currently in use to identify the neural generator(s) contributing to the scalp-recorded data (e.g., dipole modeling, minimum norm estimate, independent component analysis, and beamformer) (Picton, 2010). The various source localization methods differ in what additional assumptions are made to constrain the inverse problem. Assumptions can be made about the geometrical, anatomical, and electromagnetic properties of the brain, the spectrotemporal and/or statistical properties of the generators, and so forth. The quality of the solution largely depends on how well (1) the recorded signals cover the spatial spread and resolution of the target brain activity, (2) the target activity can be separated from other concurrent electromagnetic activity recorded in the signals (the signal-to-noise ratio), and (3) additional assumptions are met in the given situation (e.g., assuming symmetric sources reduces the complexity of the solution but at the cost of missing possible lateralized effects). A detailed discussion of the localization of electromagnetic activity in the brain can be found in Nunez and Srinivasan's book (Nunez & Srinivasan, 2006). For example, with dipole localization, the head is often modeled in first approximation as a spherical volume with layers having different electrical properties (Sarvas, 1987), and a small number of dipolar sources are assumed. Single sources in the left and right auditory cortices often sufficiently explain the distribution of auditory evoked responses across the scalp (Picton et al., 1999). For estimation of the dipole location, orientation, and strength, the difference between the measured electric field and the calculated field is minimized by varying the dipole parameters. After localizing the sources, time series of brain activity at the source location can be calculated to describe the time course of neural activity. This source space projection method can be used to transform a large array of electrodes into a smaller, more manageable number of source waveforms (Alain et al., 2009b).

The previous paragraph considered the signal averaging techniques applied in the time domain, which reveal a waveform that comprises several deflections peaking at various latencies and locations over the scalp. However, the changes in EEG following sensory stimulation can also be considered in the frequency domain. In the time-frequency analysis of the EEG, the time waveform is converted into a frequency spectrum using the discrete Fourier transform (Bertrand & Tallon-Baudry, 2000; Picton, 2010). This approach can complement the analysis in the time domain and allow researchers to examine time-locked and induced changes of EEG oscillations that occur at various frequencies (Bertrand & Tallon-Baudry, 2000; Yuval-Greenberg & Deouell, 2007). Scalp EEG oscillations can easily be observed in relation to brain states such as waking, drowsiness, and sleep. For instance, the waking state oscillations contain higher frequency signals than during deep sleep. More recent research in cognitive neuroscience revealed high frequency EEG oscillations between 25 and 80 Hz (~40 Hz; termed the gamma band) with a specific distribution that occurred during perception and cognition (Bertrand & Tallon-Baudry, 2000). These gamma oscillations may be time locked to the rate of auditory stimulation or reflect induced activity associated with perception and cognition. The latter may go unnoticed in the time domain analysis (Mazaheri & Picton, 2005) and therefore researchers should consider using both approaches while analyzing their EEG data. The amplitude of the gamma-band response is sensitive to attentional manipulation (Ross et al., 2010) and may index processes related to attentional selection and perceptual decisions. Although the exact functions of these oscillations remain a matter of active research (Mazaheri & Picton, 2005), it is generally agreed that they synchronize activity of neural networks (e.g., thalamocortical network) that may support various functions during cognition (e.g., binding of event features into an object representation).

The analysis of scalp EEG oscillations entails the estimation of time-frequency content of single trial EEG epochs using continuous wavelet transforms and the averaging of these time-frequency epochs as a function of stimulus type, task, and/or group. Figure 4.3 shows a time-frequency analysis of ERPs elicited during the oddball paradigm, which consists of presenting infrequent target sounds embedded in a sequence of homogeneous standard stimuli. These changes in rhythmic activity comprise both evoked (phase-locked) and induced (non–phase-locked) activity. The phase-locked activity is equivalent to ERPs in the time domain while the induced

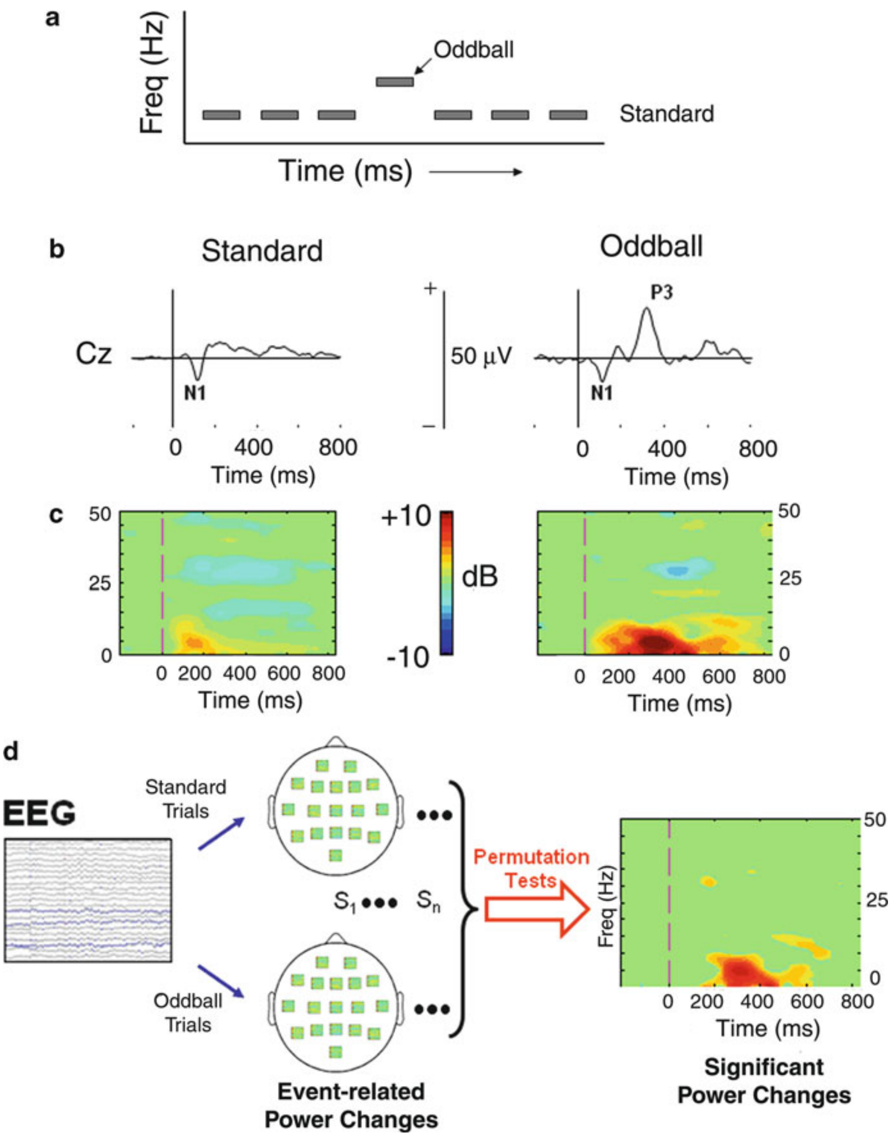


Fig. 4.3 Time-frequency analysis. (**a**) Schematic of the oddball paradigm. (**b**) Time-locked ERPs elicited by standard and target stimuli in a single participant from the midline central electrode (Cz). (**c**) Time–frequency plots that reflect averages of single-trial spectrograms for the standard and the target stimuli measured at Cz. The activity represents both evoked and induced potentials as the spectrogram of the average waveforms was not subtracted. (**d**) Schematic representation of the various steps involved in time frequency analysis

activity is best observed in the time-frequency domain. One can isolate the induced activity by subtracting the responses to a control (e.g., standard) stimulus that contains primarily transient evoked activity. The significance of these changes in oscillatory activity can then be quantified using permutation tests. Some recent research focuses on the relation between various psychological variables and the phase of these oscillations. For example, phase entrainment may mediate the effects of expectation, such as the temporal expectation of target sounds (Stefanics et al., 2010).

An important goal of ERP studies of ASA is to identify brain responses that reflect psychologically meaningful processes using task manipulations that differentially affect the latency and/or amplitude of the various deflections. Complemented with source localization, scalp recordings of ERPs can serve as a bridge between various approaches used to understand auditory perception. Recordings of ERPs are easily amenable to most perceptual and cognitive paradigms without requiring a great deal of modification. This makes them an ideal recording/imaging tool to investigate neural correlates of a particular task without introducing new variables that alter the perceptual or cognitive effects under study. ERPs also provide a means to assess perceptual and cognitive operations before and after decision-related or response-related processes because they can be recorded in conjunction with behavioral responses. More importantly, ERPs can be recorded to sounds that are not task relevant (e.g., during reading or watching a muted movie; termed the "passive" situation), thereby providing a window to explore the extent to which perceptual operations may occur outside the focus of attention. However, it should be noted that to separate the target brain responses from the electromagnetic signals resulting from other ongoing activity in the brain, the number of trials must often be increased compared to studies recording only behavioral measures.

In the following sections, research that has used ERPs to study ASA is reviewed with an emphasis on the role of prediction, attention, and learning on concurrent sound perception. This is followed by a review of the studies that have examined auditory stream segregation.

4.3 Auditory Scene Analysis as the Building Block of Higher Auditory Cognition

Current models of auditory perception recognize that hearing is more than signal detection; hearing also involves organizing the sounds that surround us in meaningful ways. This could include leisurely listening to music in a concert hall or trying to understand what someone is saying in a noisy subway station. These examples illustrate an important duality of auditory perception. On the one hand, sounds can be sorted and grouped effortlessly such that one can easily identify coherent sequences in them and possibly even localize the various sound-emitting objects; on the other hand, hearing can be effortful, requiring focused attention. Another important aspect of auditory perception is that it is inherently sensitive to the perceptual context and reflects complex interactions between data-driven grouping processes and higher-level processes that reflect knowledge and experience with the auditory environment (Bar, 2007; Alain & Bernstein, 2008; Winkler et al., 2009a). That is, incoming sounds inform our expectations of subsequent sounds. Whether listening to speech, music, or to the sound of an approaching car, we anticipate the next sounds. Such expectations play an important role for solving complex listening situations where sounds from one source may partially overlap and mask those from another source. In fact, producing predictions and generating hypotheses regarding the next sound and separating concurrent streams of sounds appear to go hand-in-hand (Winkler et al., 2009a). Prediction and attention are arguably the two processing mechanisms that enable the brain to interpret the sensory input while relying on restricted processing capacities (Summerfield & Egner, 2009).

The concept of ASA captures the dynamic nature of hearing where a mixture of sounds is perceptually organized into coherent sequences of events, termed auditory streams. Auditory streams are perceptually separable entities and their representation serves as a unit in higher-order auditory cognition. In this sense, auditory streams may be regarded as perceptual objects. For the purpose of this chapter, an auditory object refers to a mental representation rather than the actual physical sound source. It can vary from a brief sound (e.g., phone ringing) to a continuous stream (e.g., noise from the ventilation system) or a series of transient sounds (e.g., talking in the hall way). Such characterization implies that auditory objects have a particular spectrotemporal structure that appears at a particular time and place. Though this definition is fairly consistent with our subjective experience, it may lack rigor because its content varies with the perceptual context. That is, in a particular situation, a transient sound (e.g., pure tone, vowel, or phoneme) may be perceived as a distinct auditory object whereas, in another situation, it may be part of an object (e.g., melody, word). This example highlights an important property of the auditory system with respect to object formation and representation, that is, a sound may be a “part” or an “object” interchangeably depending on the context. This reflects the hierarchical nature of streams, which is especially clear in speech and music, where a word/note is embedded in a phrase, which is embedded in a sentence or longer musical passage.

How can this concept of auditory object be useful? Winkler et al. (2009a) suggested four principles, which should be applicable to all perceptual objects (for a slightly different list, see Griffiths & Warren, 2004). (1) Any notion of “object” should state that an object has discoverable features. In fact, an object is usually described by the combination of its features (e.g., Treisman, 1993). Thus, we require that auditory object representations should encode features (e.g., pitch and location) in an integrated manner (i.e., including their actual combination in the object). (2) Perceptual objects are separable from each other. That is, an object representation should allow us to decide whether a given part of the sensory input belongs to the object or not. Except for the rare case of duplex perception (e.g., Fowler & Rosenblum, 1990), the allocation of the sensory input is exclusive; that is, any part of the input belongs to exactly one object, including the borders between objects (for a discussion see Bregman, 1990). (3) Perceptual objects are nodes of invariance allowing stable interpretation of the ever-changing sensory input. In most situations, both the perceiver and some of the objects are nonstationary (e.g., they may be moving or changing their acoustic characteristics, such as the intensity of sound emission). Thus, either or both may affect the sounds arriving at the listener’s ears.

Object representations must generalize between the different ways in which the same sound source appears to our senses (e.g., transposing a melody to a different pitch). (4) Finally, sensory input usually does not contain all the information about its sources. However, common experience tells us that objects appear complete in perception, including information about parts from which no actual information has reached our senses (e.g., the back of the person standing in front of us). In line with Gregory (1980), we suggest that object representations should allow us to predict (interpolate or extrapolate) information about the object (e.g., we typically anticipate the continuation of a melody or the sound of an approaching car).

Transforming incoming acoustic data into the perception of sound objects can be assessed with ERPs and depends on a wide range of processes that include early feature extraction, grouping of acoustic elements based on their similarity and proximity, as well as higher-order cognitive and mnemonic processes that reflect our experience and knowledge of the auditory environment (Bregman, 1990; Alain & Bernstein, 2008). Many of the processes involved in ASA have possibly evolved in response to the physical regularities that are inherently present in our auditory environment. For example, acoustic energy emanating from many physical objects such as musical instruments or vocal chords encompass a fundamental frequency (F0) and (1) several harmonics that are integer multiples of the F0, (2) generally begin at the same time, (3) usually consist of smooth intensity and frequency transitions, and (4) arise from a particular location. Hence, incoming acoustic data can be grouped according to principles originally described by the Gestalt psychologists to account for visual scene analysis (Köhler, 1947), with sounds sharing the same onsets, intensities, locations, and frequencies being more likely to belong to a particular object than those that differ in these features. Other higher-level, knowledge-based, grouping processes (i.e., schema-driven) depend on experience/training during one's lifetime (i.e., we are all experts of certain types of sounds and acoustic environments, such as the sounds of our native language or those appearing in our workplace). Another important distinction refers to the memory resources required for grouping sounds that occur simultaneously (such as single notes forming a chord on a piano) and those that occur sequentially such as musical notes in a melody. The former relies on fine acoustic details in sensory memory whereas the perceptual organization of sound sequences (or streams) connects sounds separated in time and, thus, depends on memory representations describing the recent history of several auditory stimuli.

Higher-level processes such as attention play an important role in solving the scene analysis problem. First, selective attention processes can be used to "search" for weaker sound objects that might otherwise go unnoticed in adverse listening situations. Moreover, selective attention may promote audiovisual integration that in turn can facilitate perception of weak auditory signals (Helfer & Freyman, 2005; Sommers et al., 2005; Winkler et al., 2009b). Attention to a particular set of sounds (e.g., a particular talker in a crowded room) also activates schemata against which incoming acoustic data can be compared to ease sound recognition and identification. These examples emphasize the dynamic nature of ASA in which listeners capitalize on both data-driven and top-down controlled processes to generate a coherent interpretation of the sounds surrounding us.

4.4 Concurrent Sound Segregation

Scalp recordings of ERPs have proven very helpful in characterizing the psychological and neural mechanisms supporting concurrent sound perception. In the laboratory, the perception of concurrent sound objects can be induced by mistuning one spectral component (i.e., harmonic) from an otherwise periodic harmonic complex tone. Low harmonics mistuned by about 4%–6% of their original value stand out from the complex so that listeners report hearing two sounds: a complex tone and another sound with a pure tone quality (Moore et al., 1986). While the acoustic parameters that yield concurrent sound perception have been well characterized, we are only beginning to understand the neurophysiological mechanisms of this important phenomenon.

In three separate ERP experiments, Alain et al. (2001) manipulated the amount of mistuning, harmonic number, and stimulus probability while participants were either engaged in an auditory task (i.e., judging whether one or two sounds were present) or listened passively (i.e., watching a muted subtitled movie of their choice, no response required). The use of muted subtitled movies has been shown to effectively capture attention without interfering with auditory processing (Pettigrew et al., 2004). Through analysis of changes in ERP amplitude and/or latency as a function of listening condition (i.e., active vs. passive listening), inferences could be made about the timing, level of processing, and anatomical location of processes involved in concurrent sound segregation. The main finding was an increased negativity that superimposed the N1 and P2 wave elicited by the sound onset. Figure 4.4 shows the ERPs elicited by tuned and mistuned stimuli and the corresponding difference wave referred to as the object-related negativity (ORN) because its amplitude correlated with the observers' likelihood of hearing two concurrent auditory objects. The combination of EEG recording and the passive listening condition was instrumental in showing that concurrent sound segregation takes place independently of listeners' attention. The proposal that concurrent sound segregation is not under volitional control was confirmed in subsequent ERP studies using active listening paradigms that varied auditory (Alain & Izenberg, 2003) or visual attentional demands (Dyson et al., 2005).

In addition to providing evidence for primitive sound segregation, scalp recording of ERPs also revealed attention-related effects during the perception of concurrent sound objects. Indeed, when listeners are required to indicate whether they hear one or two sounds, the ORN is followed by a positive wave that peaks at about 400 ms after sound onset. This positive wave is referred to as the P400. It is present only when participants are required to make a response about the stimuli and hence is thought to index perceptual decision-making. Like the ORN, the P400 amplitude correlates with perception and is larger when participants are more likely to report hearing two concurrent sound objects. Together, these ERP studies revealed both bottom-up (attention-independent) and top-down controlled processes that are involved in concurrent sound perception.

In the ERP studies reviewed in the preceding text, the perception of concurrent sound objects and mistuning were partly confounded, making it difficult to determine whether the ORN indexes perception or the amount of mistuning. If the ORN

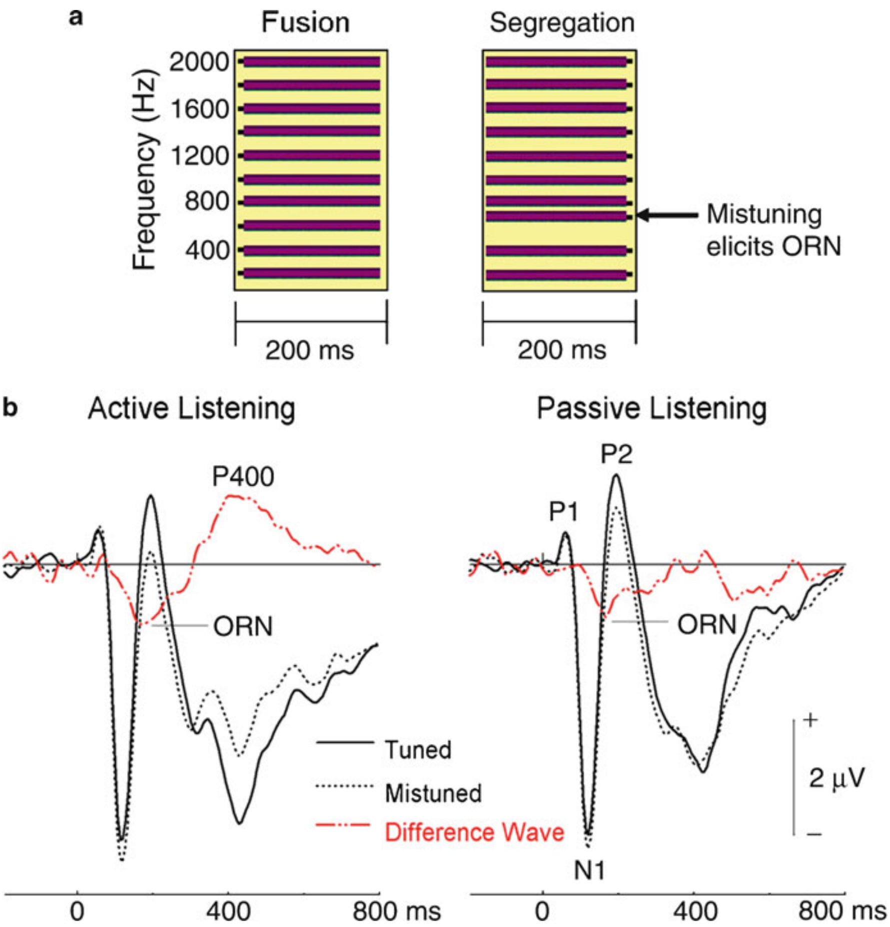


Fig. 4.4 A neural marker of concurrent sound segregation based on harmonicity. (**a**) Schematic representation of harmonic series composed of 10 pure tones with a fundamental frequency of 200 Hz. When the pure tones are all integer multiples of the fundamental, observers report hearing a buzz-like sound. However, if one of the low harmonics is mistuned by 4% or more, observers report hearing two concurrent sounds: a buzz-like sound plus another sound that has a pure tone quality. (**b**) Group mean ERPs elicited by tuned and mistuned stimuli (16% off its original value) while young healthy observers indicated whether they heard one sound object or two concurrent sound objects (Active Listening). The ERPs for the same stimuli were also recorded while participants watched a muted subtitled movie of their choice (Passive Listening). (Adapted from Alain et al., 2001.)

indexes perception of concurrent sound objects, then it should also be present when concurrent sounds are segregated on the basis of other cues such as spatial location. McDonald and Alain (2005) examined the role of location on concurrent sound perception. Using complex harmonic tones, these authors found that the likelihood of reporting that two concurrent sound objects were heard increased when a tuned or slightly mistuned harmonic was presented at a different location than the remaining harmonics. Interestingly, the effect of spatial location on perception of concurrent sound objects was paralleled by an ORN. The results from this study indicated that the ORN was not limited to mistuning, per se, but rather appeared to relate to the subjective experience of hearing two different sounds simultaneously. Moreover, this study showed that listeners can segregate sounds based on harmonicity or location alone and that a conjunction of harmonicity and location cues contributes to sound segregation primarily when harmonicity is ambiguous. This is a prime example showing that ERPs can reveal how features of a sound object may be combined during ASA.

The ERP studies reviewed in the preceding text are only a few examples of how scalp recording of ERPs can be used to investigate the brain mechanisms underlying concurrent sound segregation and perception. They show that the perceptual organization of simultaneous acoustic elements depends on the processing of fine acoustic details, which occur independently of attention. It is important to note that the changes in ERPs during concurrent sound segregation are not limited to the mistuned harmonic paradigm but have also been observed during the segregation and identification of over-learned stimuli such as speech sounds (e.g., Alain et al., 2005). Although it is not yet possible to propose a comprehensive account of how the nervous system accomplishes concurrent sound segregation, such an account will likely include multiple neurocomputational principles and multiple levels of processing in the central auditory system. To illuminate the role of concurrent sound segregation in everyday situations, future ERP studies should investigate how multiple cues (e.g., harmonicity, onset asynchrony, and spatial location) contribute to concurrent sound segregation. Does concurrent sound segregation rely on the most salient cue or does it involve a conjunction of the various cues available? Are the effects of F0 separation, spatial separation, and onset asynchrony additive or perhaps superadditive? Prior studies have shown that concurrent vowel segregation and identification can be enhanced via short-term training (Reinke et al., 2003). Would training on F0 separation generalize to segregation based on location or onset asynchrony and vice versa? These are important questions to answer if we wish to develop a more comprehensive account of ASA that applies to situations outside the laboratory. Some of these future questions may be investigated with EEG, a neuroimaging technique that will undoubtedly continue to play a pivotal role in identifying mechanisms of concurrent sound perception.

4.5 Sequential Sound Segregation

As an auditory scene unfolds over time in the real world, observers are faced with the challenge of sorting out the acoustic elements that belong to the various sound emitting objects. The critical issue is to create mental representations linking sounds separated in time. These representations must be stable, adaptable to natural variations within a stream (e.g., due to movements of the source and the listener), and capable of seamlessly absorbing the incoming sounds. This type of perceptual organization or stream segregation takes several seconds to build up with more complex scenes (e.g., in the presence of sound sources producing sounds with similar spectral features and location) requiring more time than simpler ones (e.g., sound sources that are easily distinguishable in frequency and space). In the laboratory, this form of stream segregation can be easily induced by presenting two sets of sounds differing from each other in some acoustic feature, such as in the frequency range of two sets of interleaved pure tones. In a typical paradigm, sounds are presented in patterns of "ABA—ABA—", in which "A" and "B" are tones of different frequencies and "—" is a silent interval (van Noorden, 1975). Differences in practically any sound feature can result in stream segregation (Moore & Gockel, 2002). The greater the stimulation rate and the feature separation, the more likely and more rapidly listeners report hearing two separate streams of sounds (i.e., one of A's and another of B's). Current explanations for stream segregation range from assumptions of speed limitations for attentional shift between widely different sounds (Jones et al., 1981) to neural model limitations of connecting neurons, widely separated in space, due to the tonotopic organization of the afferent auditory pathways (Hartmann & Johnson, 1991; Snyder & Alain, 2007).

Scalp recordings of ERP have proven useful in assessing these various accounts of stream segregation. Some of the ERP studies used ABA sequences similar to those shown in Figure 4.5 (e.g., Gutschalk et al., 2005; Snyder et al., 2006). These studies observed changes in sensory evoked responses that correlated with listeners' likelihood of reporting hearing two streams of sounds. Some of the changes in ERPs reflected the frequency differences between the two tones composing the sequence and were little affected by attention (Snyder et al., 2006). There was also evidence for perception-related changes in neural activity from auditory cortices independently of frequency separation (Gutschalk et al., 2005; Snyder et al., 2009b). Snyder et al. (2006) also identified an ERP modulation that mimicked the increasing likelihood of experiencing auditory streaming, which was modulated by listeners' attention. Together, these studies suggest that auditory stream segregation involves attention-independent and attention-dependent processes, and are part of an increasing effort to identify neural correlates of auditory stream formation and segregation.

The oddball paradigm and the mismatch negativity (MMN; Näätänen et al., 1978) have also been successfully employed for investigating the neural mechanisms that underlie the perceptual organization of sounds. This is because infrequent (oddball) violations in spectral and/or temporal regularities generate the MMN (Alain et al., 1999; Winkler, 2007). In the context of auditory stream segregation, the MMN can be used to make inferences about perceptual organization by varying parameters that would cause a particular stimulus to generate an MMN only if it violates spectral and/or temporal regularities in a particular stream of sounds. For example, one can hide a well known tune in a sequence by interleaving the sounds of the melody with random sounds. If, however, the random sounds can be segregated from those of the melody, the tune can be perceived again (Dowling, 1973). This phenomenon provides the grounds for using MMN in the study of auditory stream segregation (Sussman et al., 1999). By designing stimulus paradigms in which some regular feature can only be detected in one of the possible alternative organizations, the elicitation of the MMN component becomes an index of this organization. The typical stimulus paradigm and ERP results are illustrated in Figure 4.6 (Winkler et al., 2003c).

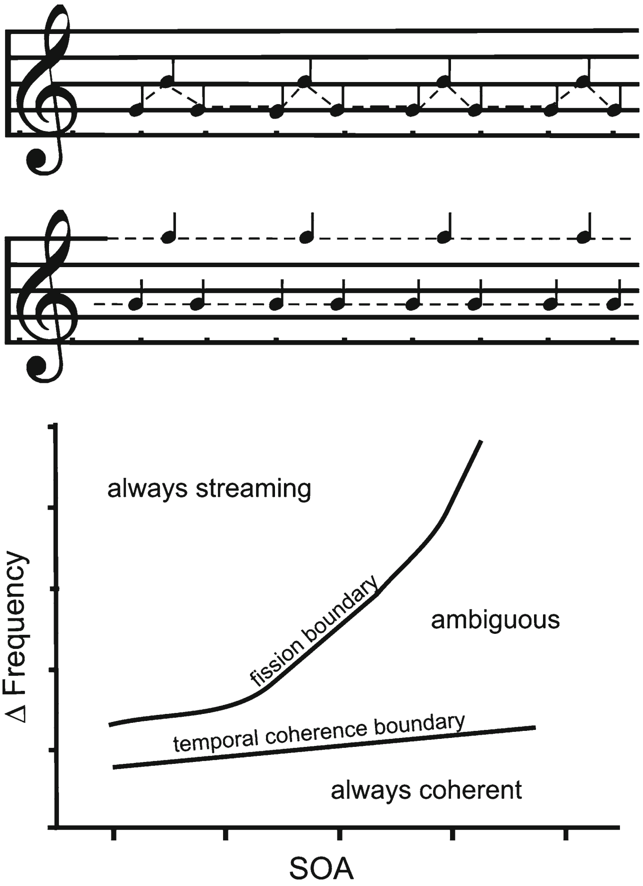


Fig. 4.5 The top part of the illustration shows a schematic diagram of the ABA- paradigm. The dashed line between the notes indicates the typical organization associated with these particular frequency separations. The bottom panel illustrates the fission and the temporal coherence boundary that can be obtained using an ABA- pattern such as the one displayed above. The fission and temporal coherence boundaries are assessed by varying the presentation rate and frequency separation between the notes and asking participants when they can no longer voluntarily hear a particular percept (i.e., one coherent sequence of tones that alternates or two streams of sounds). In other words, the incoming sequence is always heard as either one stream or two streams regardless of the frequency separation and/or stimulation rate. SOA, stimulus onset asynchrony. The functions printed in the figure are only approximations. (Adapted from Van Noorden, 1975.)

The base condition is a simple oddball sequence (Fig. 4.6, top) in which one tone is repeated most of the time (standard), occasionally replaced by a different tone (deviant). Because the deviant violates the repetition of the standard, it elicits the MMN. The MMN can be estimated from the difference between the ERP elicited by the deviant stimulus and that elicited by a similar regular (standard) stimulus (for a discussion of the optimal estimation of the MMN response, see Kujala et al., 2007).

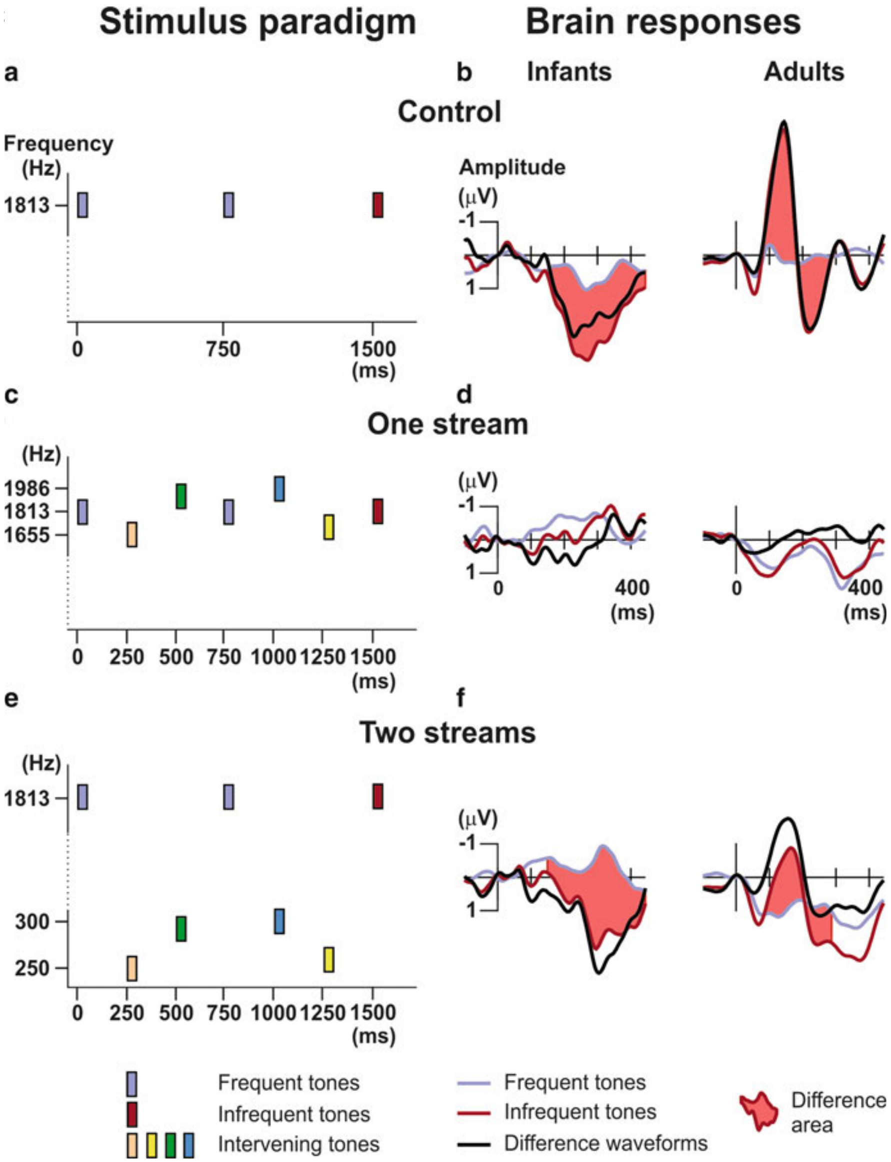


Fig. 4.6 Using MMN to test auditory streaming. (**a**) Schematic illustration of a segment of the base–condition sequence. Rectangles represent tones whose *y*-coordinate shows the tone frequency (logarithmic scale). Different loudness level settings are marked with different colors: frequent soft (standard) tones in pastel blue, infrequent louder (deviant) tones in dark red. (**b**) Frontal (F4) electric brain responses (infants, left; adults, right) elicited by the standard (pastel blue lines) and deviant tones (dark red line) together with their respective difference waveform (black line). Tone onset is at the 0 ms mark, amplitude values are calibrated in μV units. The light red shading of the area between the standard and deviant responses marks significant differences between the two brain responses. (**c**) In the single–stream condition, intervening tones varied in frequency and intensity. (**d**) The responses to the standard and deviant tones did not significantly differ from each other in either group of subjects. (**e**) For the two-streams condition, the frequencies of the intervening tones were lowered from the values used in the single–stream condition, but the tone intensity values were retained. (**f**) The responses to the standard and deviant tones significantly differed from each other in both groups of subjects and were similar to those elicited in the base condition. (Adapted from Winkler et al., 2003c.)

Inserting two additional tones between consecutive tones of the oddball sequence (Fig. 4.6, middle row) eliminates the MMN when the frequency of the intervening tones vary in a narrow range centered on the frequency of the tones of the oddball sequence. However, the MMN is elicited again when the frequencies of the intervening tones are selected from a widely different range (Fig. 4.6, bottom).

Recent computational modeling efforts (Garrido et al., 2009) suggest that both stimulus-specific neuronal adaptation (Nelken & Ulanovsky, 2007) and memory-based deviance detection processes (Winkler, 2007) contribute to the deviant-minus-standard difference waveform estimate of the MMN response. The latter process may be directly involved in ASA. Indeed, the deviance detection process underlying the generation of MMN to pattern deviant stimuli appears to rely on predictions drawn from the perceptual organization of the stimuli. That is, the MMN is elicited by violating predictive rules, such as "short tones are followed by high-pitched tones, long tones by low-pitched tones" (Paavilainen et al., 2007; Bendixen et al., 2008) and deviant sounds only elicit MMN with respect to the stream within which they belong (Ritter et al., 2000). Further, it has been shown that the primary function of the MMN-eliciting process is related to the representation of the violated regularity (Alain et al., 1994, 1999; Winkler & Czigler, 1998) as opposed to the deviant stimulus itself. Figure 4.7 shows a conceptualization of ASA and the role of the MMN-generating process in selecting the dominant sound organization.

The picture emerging from MMN studies is that auditory streams are formed and maintained by finding acoustic regularities and then using that information to generate an expectation regarding the incoming acoustic events. The initial segregation of streams is probably based on simple feature cues (such as pitch differences), whereas streams are stabilized (maintained) by finding additional, often much more complex regularities, including syntactic and semantic ones. This entails the representation of stimulus features and their conjunction (e.g., Takegata et al., 2005) as well as the representation of the spectral and temporal transitions connecting the sound elements that comprise the auditory scene (e.g., Winkler & Schröger, 1995; Alain et al., 1999). Regularities are extracted even from highly variable sequences of sounds and they can absorb natural variations of a source (Gomes et al., 1997; Alain et al., 1999; Näätänen et al., 2001). As discussed in the preceding text, the regularity representations inferred from MMN results are probably predictive. Sounds usually belong to only one regularity representation at a time (Ritter et al., 2000, 2006; Winkler et al., 2006). Thus, these representations meet the criteria suggested for (auditory) perceptual object representations.

4.6 Attention, Prediction, and Auditory Scene Analysis

There is an ongoing debate regarding the role of attention in ASA. Though, in the original formulation of the theory, the role of learning and attention was acknowledged, the primary stream segregation process was also proposed to account for our phenomenological experience. As mentioned earlier, scalp recordings of ERPs have

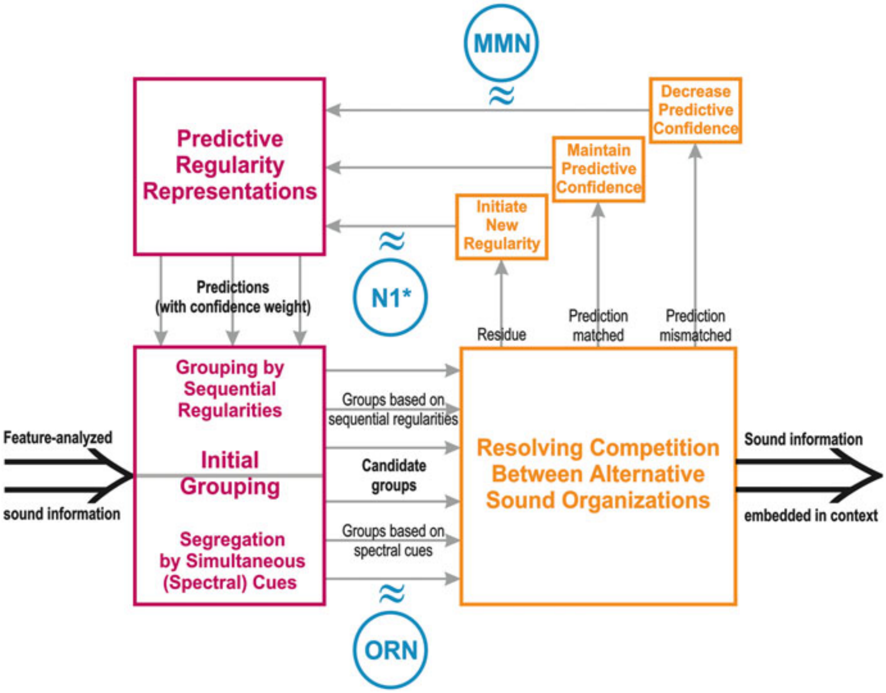

MMN
≈
Predictive Regularity Representations
Decrease Predictive Confidence
Maintain Predictive Confidence
Initiate New Regularity
≈
N1*
Predictions (with confidence weight)
Residue
Prediction matched
Prediction mismatched
Grouping by Sequential Regularities
Initial Grouping
Segregation by Simultaneous (Spectral) Cues
Feature-analyzed
sound information
Groups based on sequential regularities
Candidate groups
Groups based on spectral cues
Resolving Competition Between Alternative Sound Organizations
Sound information
embedded in context
≈
ORN

Fig. 4.7 ERP components associated with auditory scene analysis (ASA) functions. First phase of ASA (left; magenta): Auditory information enters initial grouping (lower left box); simultaneous and sequential grouping processes are marked separately. Second phase of ASA (right; orange): Competition between candidate groupings is resolved; the alternative supported by processes with the highest confidence appears in perception (lower right box). Confidence in those regularity representations (upper left box) whose predictions failed is reduced and the unpredicted part of the auditory input (residue) is parsed for new regularities (upper right boxes). Predictive regularity representations support sequential grouping (feedback to the lower left box). ERP components associated with some of the ASA functions (light blue circles linked to the corresponding function by "≈" signs): ORN reflects segregation by simultaneous cues. N1* stands for the exogenous components possibly reflecting the detection of a new stream. MMN is assumed to reflect the process of adjusting the confidence weight of those regularity representations whose predictions were not met by the actual input. (Adapted from Winkler, 2007.)

proven helpful in studying the role of attention in auditory perceptual organization because they can be recorded to sounds that are presented outside the focus of attention.

Several studies suggest that auditory streams can be maintained without attention focused on the sounds (e.g., Winkler et al., 2003a,b). However, some data are inconsistent with this view. In a series of experiments using the ABA- pattern, Carlyon and colleagues showed that the buildup in streaming is affected by intra- (Carlyon et al., 2001) and intermodal (Carlyon et al., 2003) attention; based on these results the authors proposed that attention may be needed for stream segregation to occur.

Supporting this hypothesis, patients with unilateral neglect brain damage showed impaired buildup in streaming relative to age-matched controls when stimuli were presented to the neglected side (Carlyon et al., 2001). In a subsequent study, Cusack et al. (2004) showed that switching attention from an arbitrary task to one involving the ABA- pattern disrupts auditory stream segregation and they proposed that task switching may reset the streaming process. Further support for the role of attention in auditory stream segregation was obtained by Sussman et al. (2005) with the MMN method. These authors found that when one stream was designated as task relevant, sounds from two further frequency regions were not segregated from each other, suggesting that attention is important for the perceptual organization of sounds. Lastly, there is evidence that auditory stream segregation can be modulated by presenting visual cues synchronously with the auditory stimuli (Rahne et al., 2007; Rahne & Bockmann-Barthel, 2009). Together, these studies suggest that auditory stream segregation is sensitive to top-down controlled processes.

In contrast to pro-attention evidence, there is also evidence to suggest that attention may not be required for perceptual organization to occur. For instance, Deouell et al. (2008) found that patients with unilateral neglect, although unaware of sounds presented to their neglected side, experienced the "scale illusion" (Deutsch, 1975). The scale illusion refers to the situation where observers experience one or two melodies consisting of the lower or higher portion of a scale that can only occur if the sounds from the left and right ears are grouped together. Such findings are difficult to reconcile with a model invoking a required role of attention in stream segregation and suggest that some organization must be taking place outside the focus of attention (e.g., Alain & Woods, 1994; Arnott & Alain, 2002; Sussman et al., 2007). This apparent discrepancy could be reconciled by assuming that sequential stream segregation relies on multiple levels of representation (Denham & Winkler, 2006; Snyder et al., 2009a,b), some of which may be more sensitive to perceptual context (Gutschalk et al., 2005; Snyder et al., 2006). In support of this hypothesis, Winkler et al. (2005) found ERP evidence for two distinct phases in auditory stream segregation. Presenting participants with an ambiguously segregated sequence of tones, the authors separately analyzed trials during which the participant heard one or two streams. Two successive ERP responses were found to be elicited by occasional deviants appearing in the sequence. An early (50–70 ms peak latency) negative difference between deviant and standard ERP responses was observed when stimulus parameters did not particularly promote segregation irrespective of participants' perception of the sequence in terms of one or two streams. A later (ca. 170 ms peak latency) negative difference was elicited only when participants perceived the sequence in terms of a single stream. Because both responses were related to deviance detection, these results suggest the existence of at least two different representations of the standard: one that is mainly stimulus driven and another that correlates with perception (see also Gutschalk et al., 2005; Snyder et al., 2009b).

ERP results obtained in the context of auditory stream segregation suggest that some organization of the auditory input occurs even when all or some sounds fall outside the focus of attention (e.g., Alain & Woods, 1994; Sussman et al., 2007).

However, it has been proposed that stream segregation can be reset by attention (Cusack et al., 2004) and that selective attention may determine which stream of sounds is in the foreground and which stimuli are in the background (i.e., figure–ground segmentation) (Sussman et al., 2005). These views are compatible with the suggestion that predictive processing and selective attention are two independent functions of the sensory systems in the brain with the common goal of allowing fast processing of a large amount of information using limited capacities (Summerfield & Egner, 2009). Predictive processing allows a large part of the sensory input to be absorbed at low levels of the system while maintaining a stable (Winkler et al., 1996), reasonably accurate representation of the whole environment and allowing selective processes to work on the basis of meaningful objects (Alain & Arnott, 2000). ERP studies of auditory object formation provide support for an object-based account of selective attention (e.g., Alain & Woods, 1994; Arnott & Alain, 2002; Winkler et al., 2005a, b). When novel information enters the auditory system, that is, a sound that could not be predicted, attention may be shifted to provide more in-depth processing of the new information. The process of attention-switching is thought to be reflected by the P3a wave (Alho et al., 1997; Schröger & Wolff, 1998a), which often follows the MMN but can also be elicited by isolated sounds that would not trigger an MMN. Once the attention capturing stimulus has been processed, attention can be redirected to the process that was interrupted. Reorientation of attention is accompanied by an ERP component termed the reorientation negativity (Schröger & Wolff, 1998b; Berti & Schröger, 2001).

We have already mentioned that the competition between alternative organizations can be biased by attention to a certain extent and that detection of a new sound object or a new stream can capture attention. Based on the old-plus-new principle, new streams are identified by analyzing the residual auditory signal (the part of the auditory input left unexplained by the continuation of previously detected streams). This process can be enhanced and even over-ruled by attentive processes. When actively waiting for the emergence of a sound, it is more likely that we detect it in the context of background noise and often we reinterpret the auditory scene in favor of the sound which we are expecting. Building an attentional template during selective listening is accompanied by a negative displacement of the ERP commencing during the time range of the exogenous components (Nd; see Hansen & Hillyard, 1980) and, depending on the similarity between the input and the template, possibly lasting beyond 100 ms (processing negativity [PN]; see Näätänen, 1982). Recently, Winkler et al (2009a) suggested that the exogenous auditory ERP components (P1, N1, P2) may reflect processes involved in detecting new streams. This assumption is depicted in Figure 4.7. Their timing (shortly following the initial period in which direct correlates of prediction were observed, see Bendixen et al., 2009), as well as the fact that they were elicited by sound onset, is compatible with this hypothesis. Finally, a large part of the grouping processes are schema-driven. That is, they are learned through training or exposure to certain types of sounds and acoustic environments such as the sounds of those languages that we learned to speak or the acoustic environment in our workplace. Although some of these learned grouping processes may become automatic (van Zuijen et al., 2005), others still require attention

(Carlyon et al., 2001; Snyder & Alain, 2007; Alain & Bernstein, 2008). In this way, attention can also affect what types of groups we can form and thus what sounds our auditory system will predict in a given auditory scene (this is marked in the upper left corner of Fig. 4.7).

4.7 Concluding Remarks

During the last decade, we have seen a great deal of research on ASA using ERPs. These studies reveal that perceptual organization of relatively simple sounds entails neurocomputational operations that likely include habituation and forward suppression. The ERP research highlights the importance of processing stimulus invariance and using that information to generate hypotheses regarding the incoming sound events. Despite important progress in developing and refining models of ASA, many challenges lie ahead. Most research on ASA has focused on relatively simple sounds (e.g., pure tones) and has identified a number of general principles for the grouping of sound elements (e.g., similarity, proximity). Current models of ASA are based for the most part on findings from studies using either the ABA-pattern or the mistuned harmonic paradigms, experimental situations that are well controlled but seem only vaguely relevant to real world situations. Hence, this places an important limit on how ERP studies of ASA can be used to help understand more complex and realistic listening situations often illustrated using the cocktail party example. Moreover, as currently understood, the grouping principles, and by extension current models and theories of ASA, appear inadequate to explain the perceptual grouping of speech sounds (Remez et al., 1994,) because speech has acoustic properties that are diverse and rapidly changing. Furthermore, speech is a highly familiar stimulus and so our auditory system has had the opportunity to learn about speech-specific properties (e.g., F0, formant transitions) that may assist in the successful perceptual grouping of speech stimuli (Rossi-Katz & Arehart, 2009). Lastly, spoken communication is a multimodal and highly interactive process where visual input can help listeners identify speech in noise and can also influence what is heard. Hence, it is also important to examine the role of visual information in solving complex listening situations. Do auditory object (stream) representations play a role in cross-modal integration? In face-to-face communication we understand better if we can see the speaker's lips moving (Summerfield, 1992; Benoit et al., 1994; Sommers et al., 2005). The visual information about the configuration of lips, teeth, and tongue determines the resonance of the vocal tract and conveys important phonetic aspects of speech, for example, the place of articulation of the consonants \b\ and \d\, which can ease the interpretation of acoustic information especially in adverse listening situations where many people are talking at the same time.

Other important issues that deserve further empirical research are related to identifying which cues can initiate stream segregation and which ones can stabilize the established streams. Do object (stream) representations explicitly manifest in auditory processing or only through affecting various processes? Do auditory object (stream) representations have a fixed content or are they adapted to the actual scene/context? What are the effects of development on auditory scene analysis beyond acquiring schemata? Most sound objects are defined by a combination of features and yet most of the research to date has examined streaming based on a single cue (e.g., frequency; however, see Denham et al., 2010; Du et al., 2011). That is, sounds that differ in onset asynchrony, F0, and/or spatial location are more likely to be coming from different sound objects than those that begin at the same time and share the same F0 or location. Though the contribution of each cue to concurrent sound perception has been well documented, the synergetic effect of having more than one cue available is less well understood. Many cues can contribute to sound segregation including differences in frequency, spatial location and onset asynchrony. Does sound segregation rely on the most salient cue or does it involve a conjunction of the various cues available? It appears that the effect of frequency and spatial separation are superadditive (Denham et al., 2010; however, see Du et al., 2011), but studying other feature combinations may reveal important processing principles of separating sound sources under ecologically valid circumstances. These are important questions to answer if we wish to understand how the myriad of sounds that surround us are perceptually organized in a meaningful way.

Acknowledgments We thank Stephen Arnott, Joel Snyder, and Alexandra Bendixen for their suggestions on an earlier version of this chapter, as well as Patricia Van Roon and Filomeno Cortese for their help in preparing the manuscript and illustrations.

References

Alain, C. (2007). Breaking the wave: Effects of attention and learning on concurrent sound perception. *Hearing Research*, 229(1–2), 225–236.

Alain, C., & Arnott, S. R. (2000). Selectively attending to auditory objects. *Frontiers in Biosciences*, 5, D202–212.

Alain, C., & Bernstein, L. J. (2008). From sounds to meaning: The role of attention during auditory scene analysis. *Current Opinion in Otolaryngology & Head and Neck Surgery*, 16, 485–489.

Alain, C., & Izenberg, A. (2003). Effects of attentional load on auditory scene analysis. *Journal of Cognitive Neuroscience*, 15(7), 1063–1073.

Alain, C., & Woods, D. L. (1994). Signal clustering modulates auditory cortical activity in humans. *Perception & Psychophysics*, 56(5), 501–516.

Alain, C., Woods, D. L., & Ogawa, K. H. (1994). Brain indices of automatic pattern processing. *NeuroReport*, 6(1), 140–144.

Alain, C., Cortese, F., & Picton, T. W. (1999). Event-related brain activity associated with auditory pattern processing. *NeuroReport*, 10(11), 2429–2434.

Alain, C., Arnott, S. R., & Picton, T. W. (2001). Bottom-up and top-down influences on auditory scene analysis: Evidence from event-related brain potentials. *Journal of Experimental Psychology: Human Perception and Performormance*, 27(5), 1072–1089.

Alain, C., Reinke, K., He, Y., Wang, C., & Lobaugh, N. (2005). Hearing two things at once: Neurophysiological indices of speech segregation and identification. *Journal of Cognitive Neuroscience*, 17(5), 811–818.

Alain, C., Snyder, J. S., He, Y., & Reinke, K. S. (2007). Changes in auditory cortex parallel rapid perceptual learning. *Cerebral Cortex*, 17(5), 1074–1084.

Alain, C., Quan, J., McDonald, K., & Van Roon, P. (2009a). Noise-induced increase in human auditory evoked neuromagnetic fields. *European Journal of Neuroscience*, 30(1), 132–142.

Alain, C., McDonald, K. L., Kovacevic, N., & McIntosh, A. R. (2009b). Spatiotemporal analysis of auditory what" and "where" working memory. *Cerebral Cortex*, 19(2), 305–314.

Alain, C., Campeanu, S., & Tremblay, K. (2010). Changes in sensory evoked responses coincide with rapid improvement in speech identification performance. *Journal of Cognitive Neuroscience*, 22(2), 392–403.

Alho, K., Escera, C., Diaz, R., Yago, E., & Serra, J. M. (1997). Effects of involuntary auditory attention on visual task performance and brain activity. *NeuroReport*, 8(15), 3233–3237.

Arnott, S. R., & Alain, C. (2002). Effects of perceptual context on event-related brain potentials during auditory spatial attention. *Psychophysiology*, 39(5), 625–632.

Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, 11(7), 280–289.

Bendixen, A., Prinz, W., Horváth, J., Trujillo-Barreto, N. J., & Schröger, E. (2008). Rapid extraction of auditory feature contingencies. *NeuroImage*, 41(3), 1111–1119.

Bendixen, A., Schröger, E., & Winkler, I. (2009). I heard that coming: Event-related potential evidence for stimulus-driven prediction in the auditory system. *Journal of Neuroscience*, 29(26), 8447–8451.

Benoit, C., Mohamadi, T., & Kandel, S. (1994). Effects of phonetic context on audio-visual intelligibility of French. *Journal of Speech and Hearing Research*, 37(5), 1195–1203.

Berti, S., & Schröger, E. (2001). A comparison of auditory and visual distraction effects: Behavioral and event-related indices. *Cognitive Brain Research*, 10(3), 265–273.

Bertrand, O., & Tallon-Baudry, C. (2000). Oscillatory gamma activity in humans: A possible role for object representation. *International Journal of Psychophysiology*, 38(3), 211–223.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sounds*. London: The MIT Press.

Carlyon, R. P., Cusack, R., Foxton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, 27(1), 115–127.

Carlyon, R. P., Plack, C. J., Fantini, D. A., & Cusack, R. (2003). Cross-modal and non-sensory influences on auditory streaming. *Perception*, 32(11), 1393–1402.

Cusack, R., Deeks, J., Aikman, G., & Carlyon, R. P. (2004). Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *Journal of Experimental Psychology: Human Perception and Performance*, 30(4), 643–656.

Davis, P. A. (1939). Effects of acoustic stimuli on the waking human brain. *Journal of Neurophysiology*, 2, 494–499.

Denham, S. L., & Winkler, I. (2006). The role of predictive models in the formation of auditory streams. *Journal of Physiology* (Paris), 100(1–3), 154–170.

Denham, S. L., Gyimesi, K., Stefanics, G., & Winkler, I. (2010). Stability of perceptual organisation in auditory streaming. In E. A. Lopez-Poveda, A. R. Palmer, & R. Meddis (Eds.), *The neurophysiological bases of auditory perception* (pp. 477–488). New York: Springer.

Deouell, L. Y., Deutsch, D., Scabini, D., & Knight, R. T. (2008). No disullusions in auditory extinction: Perceived a melody comprised of unperceived notes. *Frontiers in Human Neuroscience*, 1, 1–6.

Deutsch, D. (1975). Two-channel listening to musical scales. *Journal of the Acoustical Society of America*, 57(5), 1156–1160.

Dowling, W. J. (1973). Rhythmic groups and subjective chuncks in memory for melodies. *Perception & Psychophysics*, 14, 37–40.

Du, Y., He, Y., Ross, B., Bardouille, T., Wu, X., Li, L., & Alain, C. (2011). Human auditory cortex activity shows additive effects of spectral and spatial cues during speech segregation. *Cerebral Cortex*, 21(3), 698–707.

Dyson, B. J., Alain, C., & He, Y. (2005). Effects of visual attentional load on low-level auditory scene analysis. *Cognitive*, *Affective*, *& Behavioral Neuroscience*, 5(3), 319–338.

Fowler, C. A., & Rosenblum, L. D. (1990). Duplex perception: a comparison of monosyllables and slamming doors. *Journal of Experimental Psychology: Human Perception and Performance*, 16(4), 742–754.

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology*, 120(3), 453–463.

Gomes, H., Bernstein, R., Ritter, W., Vaughan, H. G., Jr., & Miller, J. (1997). Storage of feature conjunctions in transient auditory memory. *Psychophysiology*, 34(6), 712–716.

Gregory, R. L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 290(1038), 181–197.

Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nature Review Neuroscience*, 5(11), 887–892.

Gutschalk, A., Micheyl, C., Melcher, J. R., Rupp, A., Scherg, M., & Oxenham, A. J. (2005). Neuromagnetic correlates of streaming in human auditory cortex. *Journal of Neuroscience*, 25(22), 5382–5388.

Hansen, J. C., & Hillyard, S. A. (1980). Endogenous brain potentials associated with selective auditory attention. *Electroencephalography and Clinical Neurophysiology*, 49(3–4), 277–290.

Hartmann, W. M., & Johnson, D. (1991). Stream segregation and pereipheral channelling. *Music Perception*, 9, 155–184.

Helfer, K. S., & Freyman, R. L. (2005). The role of visual speech cues in reducing energetic and informational masking. *Journal of the Acoustical Society of America*, 117(2), 842–849.

Jones, M. R., Kidd, G., & Wetzel, R. (1981). Evidence for rhythmic attention. *Journal of Experimental Psychology: Human Perception and Performance*, 7(5), 1059–1073.

Köhler, W. (1947). *Gestalt psychology*. New York: Liveright.

Kujala, T., Tervaniemi, M., & Schröger, E. (2007). The mismatch negativity in cognitive and clinical neuroscience: Theoretical and methodological considerations. *Biological Psychology*, 74(1), 1–19.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.

Mazaheri, A., & Picton, T. W. (2005). EEG spectral dynamics during discrimination of auditory and visual targets. *Cognitive Brain Research*, 24(1), 81–96.

McDonald, K. L., & Alain, C. (2005). Contribution of harmonicity and location to auditory object formation in free field: Evidence from event-related brain potentials. *Journal of the Acoustical Society of America*, 118(3 Pt 1), 1593–1604.

Moore, B. C., & Gockel, H. (2002). Factors influencing sequential stream segregation. *Acta Acustica United with Acustica*, 88, 320–333.

Moore, B. C., Glasberg, B. R., & Peters, R. W. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *Journal of the Acoustical Society of America*, 80(2), 479–483.

Näätänen, R. (1982). Processing negativity: An evoked-potential reflection of selective attention. *Psychological Bulletin*, 92(3), 605–640.

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, 24(4), 375–425.

Näätänen, R., Gaillard, A. W., & Mantysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica* (Amst), 42(4), 313–329.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. (2001). "Primitive intelligence" in the auditory cortex. *Trends in Neurosciences*, 24(5), 283–288.

Nelken, I., & Ulanovsky, N. (2007). Mismatch negativity and stimulus-specific adaptation in animal models. *Journal of Psychophysiology*, 21, 221–223.

Nunez, P. L., & Srinivasan, R. (2006). *Electric fields of the brain: The neurophysics of EEG*. Oxford: Oxford University Press.

Paavilainen, P., Arajärvi, P., & Takegata, R. (2007). Preattentive detection of nonsalient contingencies between auditory features. *NeuroReport*, 18(2), 159–163.

Pettigrew, C. M., Murdoch, B. E., Ponton, C. W., Kei, J., Chenery, H. J., & Alku, P. (2004). Subtitled videos and mismatch negativity (MMN) investigations of spoken word processing. *Journal of the American Academy of Audiology*, 15(7), 469–485.

Picton, T. W. (1992). The P300 wave of the human event-related potential. *Journal of Clinical Neurophysiology*, 9(4), 456–479.

Picton, T. W. (2010). *Human auditory evoked potentials*. San Diego: Plural Publishing.

Picton, T. W., Alain, C., Woods, D. L., John, M. S., Scherg, M., Valdes-Sosa, P., et al. (1999). Intracerebral sources of human auditory-evoked potentials. *Audiology and Neurootology*, 4(2), 64–79.

Picton, T. W., Alain, C., Otten, L., Ritter, W., & Achim, A. (2000). Mismatch negativity: Different water in the same river. *Audiology and Neurootology*, 5(3–4), 111–139.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, 118(10), 2128–2148.

Rahne, T., & Bockmann-Barthel, M. (2009). Visual cues release the temporal coherence of auditory objects in auditory scene analysis. *Brain Research*, 1300, 125–134.

Rahne, T., Bockmann, M., von Specht, H., & Sussman, E. S. (2007). Visual cues can modulate integration and segregation of objects in auditory scene analysis. *Brain Research*, 1144, 127–135.

Reinke, K. S., He, Y., Wang, C., & Alain, C. (2003). Perceptual learning modulates sensory evoked response during vowel segregation. *Cognitive Brain Research*, 17(3), 781–791.

Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., & Lang, J. M. (1994). On the perceptual organization of speech. *Psychological Review*, 101(1), 129–156.

Remez, R. E., Ferro, D. F., Wissig, S. C., & Landau, C. A. (2008). Asynchrony tolerance in the perceptual organization of speech. *Psychonomic Bulletin & Reviews*, 15(4), 861–865.

Ritter, W., & Ruchkin, D. S. (1992). A review of event-related potential components discovered in the context of studying P3. *Annals of the New York Academy of Sciences*, 658, 1–32.

Ritter, W., Sussman, E., & Molholm, S. (2000). Evidence that the mismatch negativity system works on the basis of objects. *Neuroreport*, 11(1), 61–63.

Ritter, W., De Sanctis, P., Molholm, S., Javitt, D. C., & Foxe, J. J. (2006). Preattentively grouped tones do not elicit MMN with respect to each other. *Psychophysiology*, 43(5), 423–430.

Ross, B., Hillyard, S. A., & Picton, T. W. (2010). Temporal dynamics of selective attention during dichotic listening. *Cerebral Cortex*, 20(6), 1360–1371.

Rossi-Katz, J., & Arehart, K. H. (2009). Message and talker identification in older adults: effects of task, distinctiveness of the talkers' voices, and meaningfulness of the competing message. *Journal of Speech, Language, and Hearing Research*, 52(2), 435–453.

Sarvas, J. (1987). Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. *Physics in Medicine & Biology*, 32(1), 11–22.

Schröger, E., & Wolff, C. (1998a). Behavioral and electrophysiological effects of task-irrelevant sound change: A new distraction paradigm. *Cognitive Brain Research*, 7(1), 71–87.

Schröger, E., & Wolff, C. (1998b). Attentional orienting and reorienting is indicated by human event-related brain potentials. *NeuroReport*, 9(15), 3355–3358.

Snyder, J. S., & Alain, C. (2007). Toward a neurophysiological theory of auditory stream segregation. *Psychological Bulletin*, 133(5), 780–799.

Snyder, J. S., Alain, C., & Picton, T. W. (2006). Effects of attention on neuroelectric correlates of auditory stream segregation. *Journal of Cognitive Neuroscience*, 18(1), 1–13.

Snyder, J. S., Carter, O. L., Hannon, E. E., & Alain, C. (2009a). Adaptation reveals multiple levels of representation in auditory stream segregation. *Journal of Expermental Psychology: Humnan Perception and Performance*, 35(4), 1232–1244.

Snyder, J. S., Holder, W. T., Weintraub, D. M., Carter, O. L., & Alain, C. (2009b). Effects of prior stimulus and prior perception on neural correlates of auditory stream segregation. *Psychophysiology*, 46(6), 1208–1215.

Sommers, M. S., Tye-Murray, N., & Spehar, B. (2005). Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults. *Ear and Hearing*, 26(3), 263–275.

Stefanics, G., Hangya, B., Hernadi, I., Winkler, I., Lakatos, P., & Ulbert, I. (2010). Phase entrainment of human delta oscillations can mediate the effects of expectation on reaction speed. *Journal of Neuroscience*, 30(41), 13578–13585.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, 13(9), 403–409.

Summerfield, Q. (1992). Lipreading and audio-visual speech perception. *Philosophical Transaction of the Royal Society:Biological Sciences*, 335(1273), 71–78.

Sussman, E., Ritter, W., & Vaughan, H. G., Jr. (1999). An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology*, 36(1), 22–34.

Sussman, E. S., Bregman, A. S., Wang, W. J., & Khan, F. J. (2005). Attentional modulation of electrophysiological activity in auditory cortex for unattended sounds within multistream auditory environments. *Cognitive*, *Affective*, *and Behavioral Neuroscience*, 5(1), 93–110.

Sussman, E. S., Horváth, J., Winkler, I., & Orr, M. (2007). The role of attention in the formation of auditory streams. *Perception & Psychophysics*, 69(1), 136–152.

Takegata, R., Brattico, E., Tervaniemi, M., Varyagina, O., Näätänen, R., & Winkler, I. (2005). Preattentive representation of feature conjunctions for concurrent spatially distributed auditory objects. *Cognitive Brain Research*, 25(1), 169–179.

Treisman, A. I. E. (1993). The perception of features and objects. In A. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection*, *awareness*, *& control. A tribute to Donald Broadbent* (pp. 5–35). Oxford: Clarendon Press.

van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences.* Doctoral dissertation, Eindhoven University of Technology.

van Zuijen, T. L., Sussman, E., Winkler, I., Näätänen, R., & Tervaniemi, M. (2005). Auditory organization of sound sequences by a temporal or numerical regularity: A mismatch negativity study comparing musicians and non-musicians. *Cognitive Brain Research*, 23(2–3), 270–276.

Verleger, R. (1988). Event-related potentials and cognition: A critique of the context updating hypothesis and an alternative interpretation of P3. *Behavioral and Brain Sciences*, 11(3), 343–356.

Winkler, I. (2007). Interpreting the mismatch negativity (MMN). *Journal of Psychophysiology*, 21(3–4), 147–163.

Winkler, I., & Czigler, I. (1998). Mismatch negativity: Deviance detection or the maintenance of the 'standard'. *NeuroReport*, 9(17), 3809–3813.

Winkler, I., & Schröger, E. (1995). Neural representation for the temporal structure of sound patterns. *NeuroReport*, 6(4), 690–694.

Winkler, I., Karmos, G., & Näätänen, R. (1996). Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. *Brain Research*, 742(1–2), 239–252.

Winkler, I., Teder-Salejarvi, W. A., Horváth, J., Näätänen, R., & Sussman, E. (2003a). Human auditory cortex tracks task-irrelevant sound sources. *NeuroReport*, 14(16), 2053–2056.

Winkler, I., Sussman, E., Tervaniemi, M., Horváth, J., Ritter, W., & Näätänen, R. (2003b). Preattentive auditory context effects. *Cognitive. Affective*, *and Behavioral Neurosciences*, 3(1), 57–77.

Winkler, I., Kushnerenko, E., Horváth, J., Ceponiene, R., Fellman, V., Huotilainen, M., et al. (2003c). Newborn infants can organize the auditory world. *The Proceedings of the National Academy of Sciences of the USA*, 100(20), 11812–11815.

Winkler, I., Czigler, I., Sussman, E., Horváth, J., & Balazs, L. (2005a). Preattentive binding of auditory and visual stimulus features. *Journal of Cognitive Neuroscience*, 17(2), 320–339.

Winkler, I., Takegata, R., & Sussman, E. (2005b). Event-related brain potentials reveal multiple stages in the perceptual organization of sound. *Cognitive Brain Research*, 25(1), 291–299.

Winkler, I., van Zuijen, T. L., Sussman, E., Horváth, J., & Näätänen, R. (2006). Object representation in the human auditory system. *European Journal of Neuroscience*, 24(2), 625–634.

Winkler, I., Denham, S. L., & Nelken, I. (2009a). Modeling the auditory scene: Predictive regularity representations and perceptual objects. *Trends in Cognitive Sciences*, 13(12), 532–540.

Winkler, I., Horváth, J., Weisz, J., & Trejo, L. J. (2009b). Deviance detection in congruent audiovisual speech: Evidence for implicit integrated audiovisual memory representations. *Biological Psychology*, 82(3), 281–292.

Yuval-Greenberg, S., & Deouell, L. Y. (2007). What you see is not (always) what you hear: Induced gamma band responses reflect cross-modal interactions in familiar object recognition. *Journal of Neuroscience*, 27(5), 1090–1096.

Chapter 5
Magnetoencephalography

Srikantan Nagarajan, Rodney A. Gabriel, and Alexander Herman

5.1 The Case for Magnetoencephalography Imaging

Multiple modalities of noninvasive functional brain imaging have made a tremendous impact in improving our understanding of human auditory cortex. Since its advent in 1991, functional magnetic resonance imaging (fMRI) has emerged as the predominant modality for imaging of the functioning brain, for several reasons. As discussed by Talavage and Johnsrude (Chapter 6), fMRI uses MRI to measure changes in blood oxygenation level–dependent (BOLD) signals due to neuronal activation. It is a safe, noninvasive method that allows for whole-brain coverage, including the ability to examine activity in deep brain structures. Importantly, the widespread availability of commercial and open-source tools for analysis of fMRI data has enabled many researchers to easily embrace this technology. However, because the BOLD signal is only an indirect measure of neural activity and is fundamentally limited by the rate of oxygen consumption and subsequent blood flow mechanism, fMRI lacks the temporal resolution required to image the dynamic and oscillatory spatiotemporal patterns that are associated with cognitive processes. The temporal resolution limitations of fMRI particularly constrain auditory studies because auditory stimuli and responses have inherently fast dynamics that cannot be readily assessed with fMRI. Further, because the BOLD signal is only an approximate, indirect measure of neural activity, it might not accurately reflect true neuronal processes especially in regions of altered vasculature. In fact the exact frequency-band of neuronal processes that corresponds to the BOLD signal is still being actively debated (Logothetis et al., 2001; Niessing et al., 2005). Finally, in the context of auditory studies, because fMRI measurements involve loud scans, caused by fast

S. Nagarajan (✉) • R.A. Gabriel • A. Herman
Department of Radiology and Biomedical Imaging, University of California,
San Francisco, 513 Parnassus Avenue, S362, San Francisco, CA 94143, USA
e-mail: sri@ucsf.edu; rodney.gabriel@ucsf.edu; alexander.herman@ucsf.edu

D. Poeppel et al. (eds.), *The Human Auditory Cortex*, Springer Handbook
of Auditory Research 43, DOI 10.1007/978-1-4614-2314-0_5,

forces on MR gradient coils, the scans themselves will invoke auditory responses that have to be deconvolved from the signals in order to examine external stimulus related activity. Hence, to image brain activity noninvasively on a neurophysiologically relevant timescale and to observe neurophysiological processes more directly, silent imaging techniques are needed that have both high temporal and adequate spatial resolution.

Temporal changes, especially relating to auditory cortical function, can be noninvasively measured using methods with high (e.g., millisecond) temporal resolution, namely magnetoencephalography (MEG) and electroencephalography (EEG). MEG measures tiny magnetic fields outside of the head that are generated by neural activity. EEG is the measurement of electric potentials generated by neural activity using an electrode array placed directly on the scalp (see Alain and Winkler, Chapter 4). In contrast to fMRI, both MEG and EEG directly measure electromagnetic (EM) fields emanating from the brain with excellent temporal resolution (<1 ms) and allow the study of neural oscillatory processes over a wide frequency range (≥1–600 Hz). MEG and EEG also provide complementary information about brain activity because of their differing sensitivity to current sources within the brain. Whereas MEG is primarily sensitive to tangential currents in the brain closer to the surface and insensitive to poor conductive properties of the skull, EEG is primarily sensitive to radial sources while being highly sensitive to the conductive properties of the brain, skull, and scalp. Because bioelectric currents produced by neurons also generate magnetic fields, which are not distorted by the heterogeneous environment, measurements of these magnetic fields using MEG can be considered to give rise to an undistorted signature of underlying cortical activity. Therefore, MEG and EEG can be viewed as being complementary in terms of the sensitivity to underlying neural activity.

In this chapter, a review is initially presented on how brain activity can be reconstructed from MEG measurements with implications for spatial and temporal resolution of such reconstructions. Subsequently, a review of auditory neuroscience studies in humans that have used MEG is presented.

5.2 Sensing the Brain’s Magnetic Fields

Biomagnetic fields detected by MEG are extremely small, in the tens-to-hundreds of femto-Tesla (fT) range—seven orders of magnitude smaller than Earth’s magnetic field, and as a result, appropriate data collection necessitates a magnetically shielded room and highly sensitive detectors—superconducting quantum interference devices (SQUIDs). The fortuitous anatomical arrangement of cortical pyramidal cells allows the noninvasive detection of their activity by MEG. The long apical dendrites of these cells are arranged perpendicularly to the cortical surface and parallel to each other, allowing their electromagnetic fields to often sum up to magnitudes large enough to detect at the scalp. Synchronously fluctuating dendritic currents result in electric and magnetic dipoles that produce these electromagnetic fields (Nunez & Srinivasan, 2006). These dendritic currents from the brain are typically sensed using detection coils called flux transformers or magnetometers, which are positioned closely to the scalp and connected to SQUIDS. SQUIDS act as a magnetic-field-to-voltage converter, and its typically nonlinear response is linearized by flux-locked loop electronic circuits, and have a sensitivity of ~10 femto-Tesla per square root of Hz which is adequate for detection of brain's magnetic fields (Vrba & Robinson, 2002).

MEG sensors are often configured for differential magnetic field measurements to reduce ambient noise in measurements—which are also referred to as gradiometers, although some MEG systems are also built out of magnetometers and rely on magnetic shielding and clever electronics for noise cancellation. The two commonly used gradiometer configurations are axial and planar gradiometers. Axial gradiometers consist of two coils that share an axis, whereas planar gradiometers measure gradients (or differences) of magnetic fields in a given plane. The sensitivity profile of planar gradiometer sensors is somewhat similar to EEG, in the sense that a sensor is maximally sensitive to a source closest on the cortical surface to it. However, the sensitivity profile of an axial gradiometer can be somewhat counterintuitive because it is not maximally sensitive to sources closest to the sensors. Further, both planar and axial gradiometers are sensitive to the orientation of the sources in a counterintuitive manner, similar to EEG sensors.

Modern MEG systems often consist of simultaneous recordings from many differential sensors that cover the whole head, and total number of sensors varies from 100 to 300. The advent of such array systems has significantly advanced MEG studies. Typical MEG systems have sensors that are spaced approximately 2.2–3.6 cm apart. Although the maximum sampling rate for many MEG systems is approximately 12 kHz, most MEG data are usually recorded at about 1000 Hz, thereby still providing excellent temporal resolution for measuring the dynamics of cortical neuronal activity at the millisecond level.

The majority of auditory studies published to date using MEG have used it mainly as an electrophysiological assay of auditory sensitive brain regions. These studies focus on response properties of specific sensors within an array of sensors (or sometimes spatial averages of specific groups of sensors) and examine component peaks in sensor waveforms. Figure 5.1A shows a typical sensor configuration and magnetic field sensor responses to simple auditory stimuli. Figure 5.1B shows typical position of magnetic field sensors relative to the head and brain surfaces. Figure 5.1C overlay shows the topographic layout of the magnetic field response recorded at 100 ms and 200 ms after the onset of an acoustic stimulus.

There are many reasons why auditory neuroscientists have embraced MEG. First, MEG setup time is very short and convenient for both experimenters and subjects. A participant or patient can be in the scanner within 10–15 minutes from entering the laboratory because—unlike EEG—the lengthy time necessary to apply and check electrodes is obviated. Second, the anatomical location of large parts of auditory cortex in the human brain in the lateral sulcus makes MEG ideally suited for electrophysiological studies in audition. Further, with whole-head sensor arrays, MEG is also well suited to investigate hemispheric lateralization effects based on

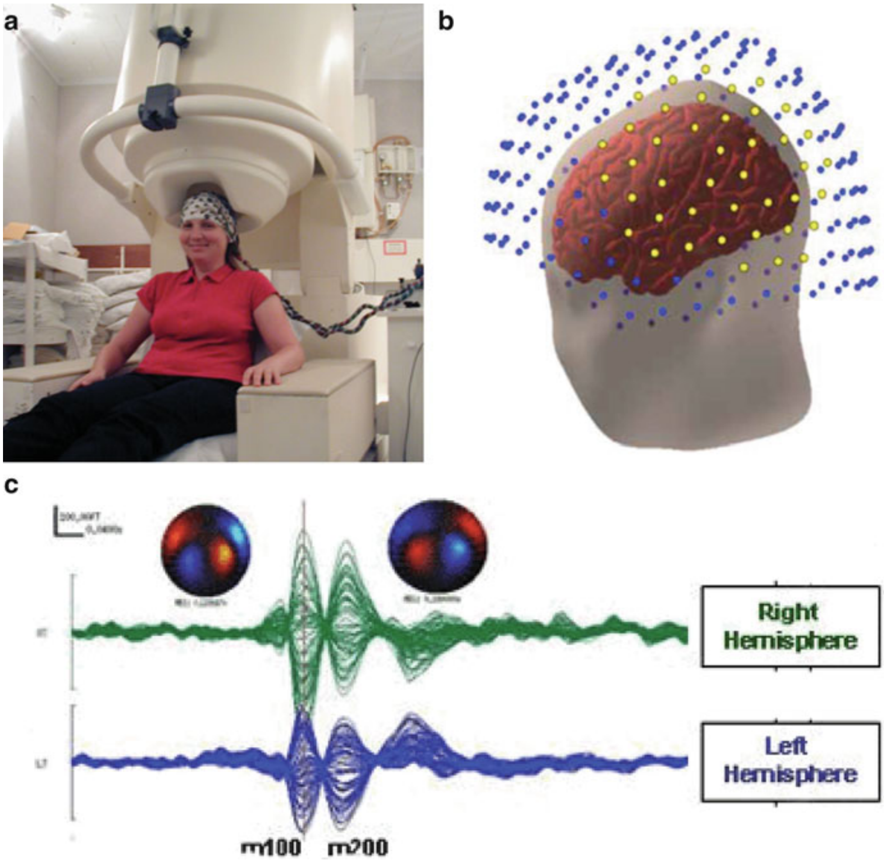


Fig. 5.1 (**a**) A subject seated just below a whole-head MEG sensor array. The subject is shown wearing a high-density electrode cap for optional simultaneous recording of EEG along with MEG. MEG measurements can be obtained by raising the seat such that the subject's head is closer to the MEG sensors located inside the beige cylindrical structure, also referred to as the dewar. The dewar is filled with liquid helium and contains superconducting quantum interference devices (SQUIDs), which are the magnetic field sensors. (**b)** Positioning of the head within the MEG sensor array. Left-hemisphere sensor locations often submitted to root-mean-square analysis are highlighted in yellow to provide an example of sensors that provide broad coverage over the auditory cortical regions. An analogous set of sensors covering the right hemisphere are also often used for hemispheric comparisons. **(c**) Auditory evoked magnetic field responses of right (green) and left hemisphere (blue) sensors to a single 400 ms long 1-kHz tone. A small-amplitude early response peak can be seen around 50 ms in the right hemisphere, followed by a dominant response peak around 100 ms, called the M100 or N100m peak. For this subject, two additional prominent peaks can also be observed around 200 ms and around 300 ms, although these peaks are not common in all subjects. The inset colored circles above the waveforms are topographic plots of the magnetic field profile at latencies corresponding to the M100 and M200 responses. This "butterfly" shaped magnetic field pattern suggests two sources located in each auditory cortex can account for these response peaks

sensor waveforms. In contrast to evoked responses measured with EEG, which are maximal at midline electrodes and therefore making hemispheric effects difficult to characterize, MEG responses are well lateralized. Distinct groups of MEG sensors are sensitive to lateralized temporal lobe activity that allows for hemisphere specific assessments.

5.3 From Sensing to Imaging

MEG sensor data analysis only provides qualitative information about underlying brain regions whose activity is observed on the sensor array based on experienced users' intuitions about the sensitivity profile of the sensors. To interpret observed sensor data more precisely in terms of the underlying brain activity, it is possible to reconstruct brain activity from MEG data. Reconstruction of brain activity from MEG data typically involves two major components: a forward model and an inverse model.

5.3.1 Forward Models Describing Brain Activity and Measurements

The forward model consists of three subcomponents: a source model, a volume conductor, and a measurement model. Typical source models assume that the MEG measurements outside the head are generated primarily by electric current dipoles located in the brain. This model is consistent with available measurements of coherent synaptic and intracellular currents in cortical columns that are thought to be major contributors to MEG and EEG signals. Although several more complex source models have been proposed recently, the equivalent current dipole is still the dominant source model in the literature. Given the distance between the sources in the brain and the sensors outside the head, the dipole is still a reasonable approximation of the sources.

Volume conductor models refer to the equations that govern the relation between the source model and the sensor measurements, that is, the electric potentials or the magnetic fields. These surface integral equations, obtained by solving Maxwell's equations under quasi-static conditions, can be solved analytically for special geometries of the volume conductor, such as a sphere and ellipsoids. For realistic volume conductors, various numerical techniques such as finite-element and boundary-element methods are employed. These methods are very time consuming and their use may appear impractical in many settings because of the lack of knowledge about specific parameters used in these models (Mosher et al., 1999a).

Measurement models refer to the specific measurement systems used in EEG and MEG including the position of the sensors relative to the head. For instance, different MEG systems measure axial versus planar gradients of the magnetic fields with respect to different location of reference sensors. The measurement model incorporates such information about the type of measurement and the geometry of the reference sensors. Because MEG sensor arrays are fixed relative to the head of a subject, it is necessary to measure the position of head relative to the sensor array. Typically this is accomplished by attaching head-localization coils to fiducial landmarks on the scalp, passing current through these coils, measuring the magnetic field created by the currents passed, and triangulating to locate the head-position relative to the sensor array. In many MEG systems, head localization is accomplished every 5–10 minutes because it disrupts normal data collection. Within a block of 10 minutes, with subjects in a supine position with their heads securely positioned in the array, typically head movements are found to be less than 5 mm. However, more modern systems are sometimes equipped with continuous head-localization procedures that enable constantly updating the sensor locations relative to the head and also correcting for subjects head movements.

The source, volume conductor and measurement models are typically combined and embodied in the idea called the "forward-field" that describes a linear relationship between sources and the measurements. Usually, we assume that the forward-field matrix is known. We can easily calculate the forward field for equivalent electric current dipoles in a spherical volume conductor model for a whole-head axial gradiometer MEG system. In this model, MEG is sensitive only to the tangential component of the primary current dipoles, whereas EEG is sensitive to all components but sensitive to uncertainties in the head model. Simultaneous MEG and EEG can be acquired in most modern MEG systems and require some modification to the forward-field matrix for combined MEG/EEG measurements, especially for more realistic source, volume conductor, and measurement models.

Coregistration is an integral part of forward model construction. Coregistration involves defining three fiducial points on an individual subject's head surface, which creates the x, y, z coordinate system that includes the brain and the position of the MEG sensors relative to it. Based on these fiducial landmarks, a transformation matrix is obtained that enables coregistration with the subjects MRI. This allows for the source locations and sensors to be defined in MRI coordinates and enables interpretation of inverse model reconstructions in terms of the underlying brain anatomy provided by MRI.

5.3.2 Inverse Models for Reconstructing Brain Activity from Measurements

Inverse algorithms are used to solve the bioelectromagnetic inverse problem, that is, estimating neural source model parameters from MEG and EEG measurements obtained outside the human head. Because the source distributions are inherently four-dimensional (three in space and one in time) and only a few measurements are made outside the head, estimation is ill posed, in other words there are no unique solutions for a given set of measurements. To circumvent this problem of non-uniqueness, various estimation procedures incorporate prior knowledge and constraints about

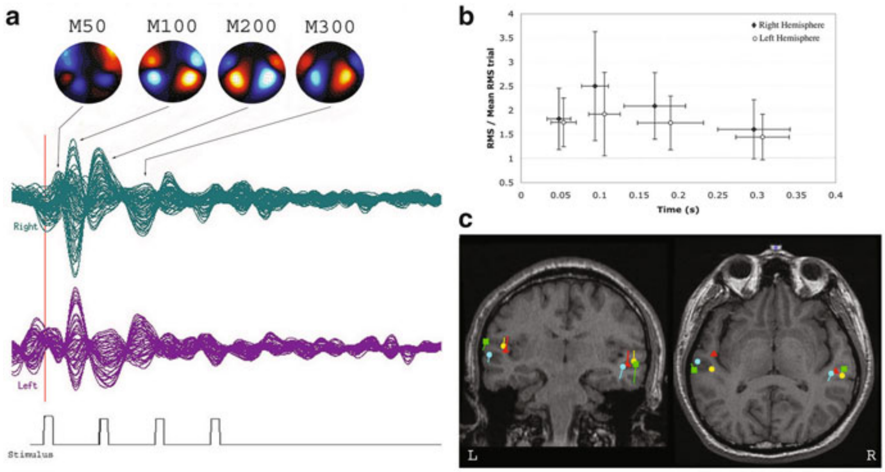


Fig. 5.2 (**a**) Auditory evoked responses to a train of tone pips occurring 200 ms apart. Blue waveforms correspond to the right hemisphere and the purple waveforms correspond to the left hemisphere. The magnetic field topography on the sensor array is shown as colored circles above for the first four peak responses. (**b**) Amplitude and latencies of the first four response peaks showing hemispheric similarities in latency and amplitudes. (**c**) Dipole localization of each of the four peaks shows activity arising from auditory cortex and its immediate environs. (Adapted from Hairston & Nagarajan, 2007.)

source characteristics such as possible source locations, the source spatial extent, the total number of sources or the source frequency/time-frequency characteristics.

Inverse algorithms can be broadly classified into two categories: parametric dipole fitting and tomographic imaging methods. Parametric dipole fitting methods assume that a small set of current dipoles (usually 2–5) can adequately represent some unknown source distribution. In this case, the dipole locations and moments form a set of unknown parameters that are typically found using either a nonlinear least square fit or multiple signal classification algorithms (MUSIC) or maximum likelihood estimation methods (Mosher et al., 1999b). Parametric dipole fitting has been successfully used clinically for localization of early sensory responses in somatosensory and auditory cortices. Figure 5.2 shows an example of parametric dipole localization in the context of auditory evoked responses, and shows that responses to early auditory peaks can often be localized to activity arising from source located in the superior temporal plane, from auditory cortex and its immediate environs. However, the localization of higher order auditory cortical functions is not always consistent and reliable with these methods either across paradigms or across subjects.

Two major problems exist in dipole fitting procedures. First, due to nonlinear optimization there are problems of local minima when more than two dipole parameters are estimated and this is usually manifested by sensitivity to initialization and some subjectivity is involved in evaluating the validity of solutions. Brute-force search methods have a huge computational burden—exponential in the number of parameters. A second, more difficult, problem in parametric methods is that often these methods require a priori knowledge of the number of dipoles. Often, such information about model order is not known a priori, especially for complex brain mapping conditions. Although information and decision theoretic criteria have been proposed to address this problem, the success of these approaches is currently unclear, especially in real data sets. Although parametric dipole methods are ideal for point or focal sources, they perform poorly for distributed clusters of sources. Nevertheless, many auditory studies to date using MEG have used dipole-fitting procedures to make inferences about auditory cortical activity.

Tomographic imaging is an alternative approach to the inverse problem. These methods impose constraints on source locations, based on anatomical and physiological information that can be derived from information obtained with other imaging modalities. Anatomical MRI provides excellent spatial resolution of head and brain anatomy, whereas fMRI techniques provide an alternative measure of neural activation based on associated hemodynamic changes. Because of the high degree of overlap in activity measured using multiple modalities, such information can be used to improve solutions to the inverse problem. If we assume that the dominant sources are the transmembrane and intracellular currents in the apical dendrites of the cortical pyramidal cells, the source image can, therefore, be constrained to the cortex, which can be extracted from a registered volume magnetic resonance image of the subjects' head. Further, the orientation of the cells normal to the cortical surface can be used to constrain the orientation of the cortical current sources. By tessellating the cortex into disjoint regions and representing sources in each region by an equivalent current dipole oriented normal to the surface, the forward model relating the sources and the measurements can be written as a linear model with additive noise. Such a formulation transforms the inverse problem into a linear imaging method because it now involves the estimation of electrical activity at discrete locations over a finely sampled reconstruction grid based on discrete measurements. This imaging problem, although linear, is also highly ill posed because of the limited number of sensor measurements available in comparison to the number of elements used in the tesselation grid.

Various solutions have been proposed for solving the tomographic imaging problem, and because there are many more unknowns to simultaneously estimate (source amplitude and time courses) than there are sensor data, the problem is therefore underdetermined.

Instead of simultaneous estimation of all sources a popular alternative is to scan the brain and estimate source amplitude at each source location independently. It can be shown that such scanning methods are closely related to whole-brain tomographic methods, and the most popular scanning algorithms are adaptive spatial filtering techniques, more commonly referred to as "adaptive beamformers" or just "beamformers" (Sekihara & Nagarajan, 2008).

Adaptive beamformers have been shown to be quite simple to implement and are powerful techniques for characterizing cortical oscillations and are closely related to other tomographic imaging methods. However, one major problem with adaptive beamformers is that they are extremely sensitive to the presence of strongly correlated sources. Although they are robust to moderate correlations, in the case of auditory studies, because auditory cortices are largely synchronous in their activity across the two hemisphere, these algorithms tend to perform poor for auditory evoked data sets without workarounds (see Fig. 5.5), and many modifications have been proposed for reducing the influence of correlated sources (Dalal et al., 2006). The simplest such workaround is to use half the sensors corresponding to each hemisphere separately, and this approach works surprisingly well for cross-hemispheric interactions. Other modifications to the original algorithms have been proposed in the literature that require some knowledge about the location of the correlated source region (Dalal et al., 2006; Quraan et al., 2010).

Many algorithms have also been proposed for simultaneous estimation of all source amplitudes, and such solutions require specification of prior knowledge about the sources either implicitly or explicitly specified in the form of probability distributions, and in these cases the solutions often require a Bayesian inference procedure of estimating some aspect of the posterior distribution given the data and the priors. Recently, we showed the many seemingly disparate algorithms for tomographic source imaging can be unified and shown, in some cases to be equivalent, using a hierarchical Bayesian modeling framework with a general form of prior distribution (called Gaussian scale mixture) and two different types of inferential procedures (Wipf & Nagarajan, 2008). These insights allow for continued development of novel algorithms for tomographic imaging in relation to prior efforts in this enterprise. Recent algorithms have shown that significant improvements in performance can be achieved by modern Bayesian inference methods that allow for accurate reconstructions of a large number of sources from typical configurations of MEG sensors (Zumer et al., 2007, 2008; Wipf et al., 2010). Figure 5.3 shows source reconstructions of auditory evoked responses using one such novel algorithm, as well as reconstructions from popular benchmark algorithms for comparisons that highlight their poorer spatial resolution and sensitivity to correlated sources and noise.

5.3.3 Sources of Noise in MEG

Even though significant breakthroughs have occurred in the source reconstruction algorithm development effort, an enduring problem in MEG and EEG based imaging is that the brain responses to sensory or cognitive events is small when compared to the large number of sources of noise, artifacts (biological and nonbiological), and interference from spontaneous brain activity unrelated to the sensory or cognitive task of interest. All existing methods for brain source localization are hampered by these many sources of noise present in MEG/EEG data. The magnitude of the stimulus-evoked auditory cortical sources are on the order of noise on a single trial, and so typically 75–200 averaged trials are at least needed to clearly distinguish the sources above noise. This limits the type of questions that can be asked, and is prohibitive for examining processes such as learning that can occur over just one or several trials.

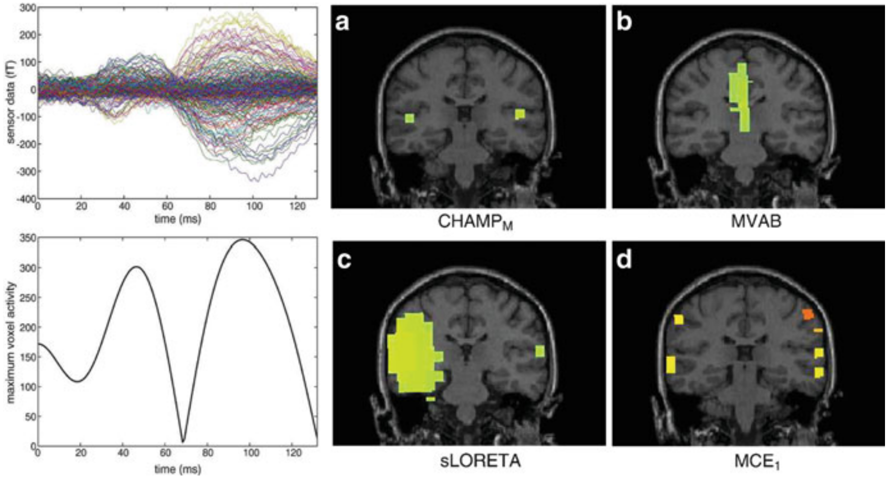


Fig. 5.3 (**a**) Top row: Auditory evoked field sensor data for a single subject showing the first 130 ms response. Bottom row: Reconstructed time course of activity in both auditory cortices. (**b**) Reconstructions of activity in this entire time window using four different source reconstruction algorithms. Top left shows reconstructions from Champagne, a recently published algorithm, that shows bilateral auditory cortex activation. Top right shows failed reconstructions from standard minimum-variance adaptive beamformers highlighting their sensitivity to strongly correlated sources. Bottom left shows reconstructions from a variant of the minimum-norm method called sLORETA which shows blurred reconstructions. Bottom right shows reconstruction from a variant of MCE, a sparse reconstruction method that is not always reliable for reconstructing distributed sources. (From Wipf et al., 2010.)

Needing to average trials is time consuming and therefore difficult for a subject or patient to hold still or pay attention through the duration of the experiment. Gaussian thermal noise or Gaussian electrical noise is present at the MEG or EEG sensors themselves. Background room interference such as from power lines and electronic equipment can be problematic. Biological noise such as heartbeat, eye blink, or other muscle artifact can also be present. Ongoing brain activity itself, including the drowsy-state alpha (~10 Hz) rhythm, can drown out evoked brain sources.

Noise in MEG and EEG data is typically reduced by a variety of preprocessing algorithms before being used by source localization algorithms. Simple forms of preprocessing include filtering out frequency bands not containing a brain signal of interest. In addition and more recently, independent component analysis (ICA) has been used to remove artifactual components, such as eye blinks (Makeig et al., 1997; Delorme & Makeig, 2004). More sophisticated techniques have also recently been developed using graphical models for preprocessing before source localization (Nagarajan et al., 2006, 2007). Therefore, algorithms for source localization from MEG and EEG data typically use a two-stage procedure—the first for noise/interference removal and the second for source localization. However, more recent algorithms that integrate interference suppression with source reconstructions have also been proposed and provide for robust source reconstruction (Zumer et al., 2007; Wipf et al., 2010).

5.3.4 Temporal and Spatial Resolution of MEG Imaging

Because MEG data can be acquired at a submillisecond time scale, temporal resolution of MEG imaging is limited only by the sampling rate, typically approximately 1 kHz, and in principle, cortical oscillations can be observed up to 500 Hz. In contrast to its temporal resolution, determining the spatial resolution of MEG imaging has been challenging because it is highly dependent on the reconstruction algorithm chosen, as well as variety of factors such as signal-to-noise and interference ratios, model formulation, forward-model accuracy, coregistration errors, and accuracy of priors. In general, it can be easily shown that the spatial resolution of MEG reconstruction is not limited by sensor spacing, because many adaptive methods can perform better than estimates based on spatial sampling criteria. For instance, while sensor spacing in many axial gradiometer systems is 2.2 cm, reconstruction accuracy can in some cases be as small as 3 mm! In general, coregistration errors alone can account for about a 3 mm accuracy in localization information for dipole fitting procedures. Whereas tomographic imaging algorithms, such as minimum-norm methods, have poor spatial resolution on the order of a few centimeters, the spatial resolution of adaptive spatial filtering methods and more recent tomographic reconstruction methods based on machine learning techniques is difficult to generally compute because these estimates depend on the data and factors contributing to data quality. As a rule of thumb, for typical data sets, these newer methods can reconstruct tens-to-hundreds of sources about 0.5 cm apart (assuming time-frequency separation and detectability) and this can be considered an approximate spatial resolution for MEG, keeping in mind that under certain circumstances the spatial resolution can be even greater.

A common myth, related to the spatial resolution of MEG, is its lack of sensitivity to gyral crown activity and relative insensitivity to deep sources. Although it is a fact that for single spherical volume conductor models MEG sensors are insensitive to radially pointing dipoles, this does not necessarily translate to gyral sources. It has been shown that, using realistic volume conductor models (such as boundary element methods or multiple local-sphere models), some sensitivity to radial sources can be recovered, and that there is no predominant loss of sensitivity to gyral sources (Hillebrand & Barnes, 2002). Further, while there is a significant drop in sensitivity to deeper sources because their contributions will fall by approximately the square of the distance to the sensors, recovery of deep sources is an issue of the signal-to-noise ratio. In general, if high signal-to-noise ratio (SNR) data are recorded, there is no inherent problem in recovery of deep sources with some of the newer Bayesian reconstruction methods. However, mid-brain sources have two additional problems. First, they may not have dipolar organization because of the architecture, although dipole approximation may not be inaccurate given the distance to the sensors the uncertainties in the lead-field increase for deep brain sources, making them more difficult to reconstruct.

5.3.5 From Single-Subject Reconstructions to Group-Level Inference

Although the power of MEG imaging is its ability to reconstruct accurately the timing of activation across different frequency bands in single subjects (see Fig. 5.4), inferences across subjects require group-level statistical analyses (Dalal et al., 2008). The most ubiquitous forms of group analysis of MEG studies of auditory cortex are based on parameters, obtained from dipole fitting of typical component peaks in the response, such as timing, amplitude, location, and sometimes orientation. For the less common tomographic and scanning based algorithms, group analyses of data across subjects have typically paralleled similar procedures for whole-brain analysis based on fMRI and positron emission tomography (PET) studies (Singh et al., 2002, 2003). These procedures include spatial normalization to template brains, general-linear modeling of experimental effects, parametric and nonparametric inference procedures, and corrections for multiple comparisons. It is to be noted that group-level statistical corrections for multiple comparisons are not yet as well developed for MEG imaging studies as they are for fMRI, and fMRI correction procedures such as family-wise errors (FWE) can sometimes be too conservative for MEG reconstructions for a variety of reasons, including the fact that spatial correlations in reconstructed images are higher than in fMRI (Darvas et al., 2004; Dalal et al., 2008).

5.4 Auditory Studies Using MEG

Numerous studies have used MEG to characterize the responses to different types of acoustic stimuli, ranging from nonspeech tones, to elemental speech sounds such as vowels and syllables, to complex speech sounds including words and sentences. MEG studies focusing on each of these stimulus categories are discussed later. Although many advances have emerged in source reconstructions from MEG data, as mentioned earlier, many MEG studies focus on timing and morphology of sensor measurements and perform only rudimentary source analyses, such as dipole fitting. Commonly, studies focus on analysis of latency and amplitude of individual response peaks. Cortical activation sequences are often characterized by examining the location of different activation peaks. More recently, cortical oscillations induced by auditory stimuli have been studied (Palva et al., 2002), including studies examining the phase relationship between acoustic stimuli and the MEG measurements (Ahissar et al., 2001; Patel & Balaban, 2004; Ross et al., 2007).

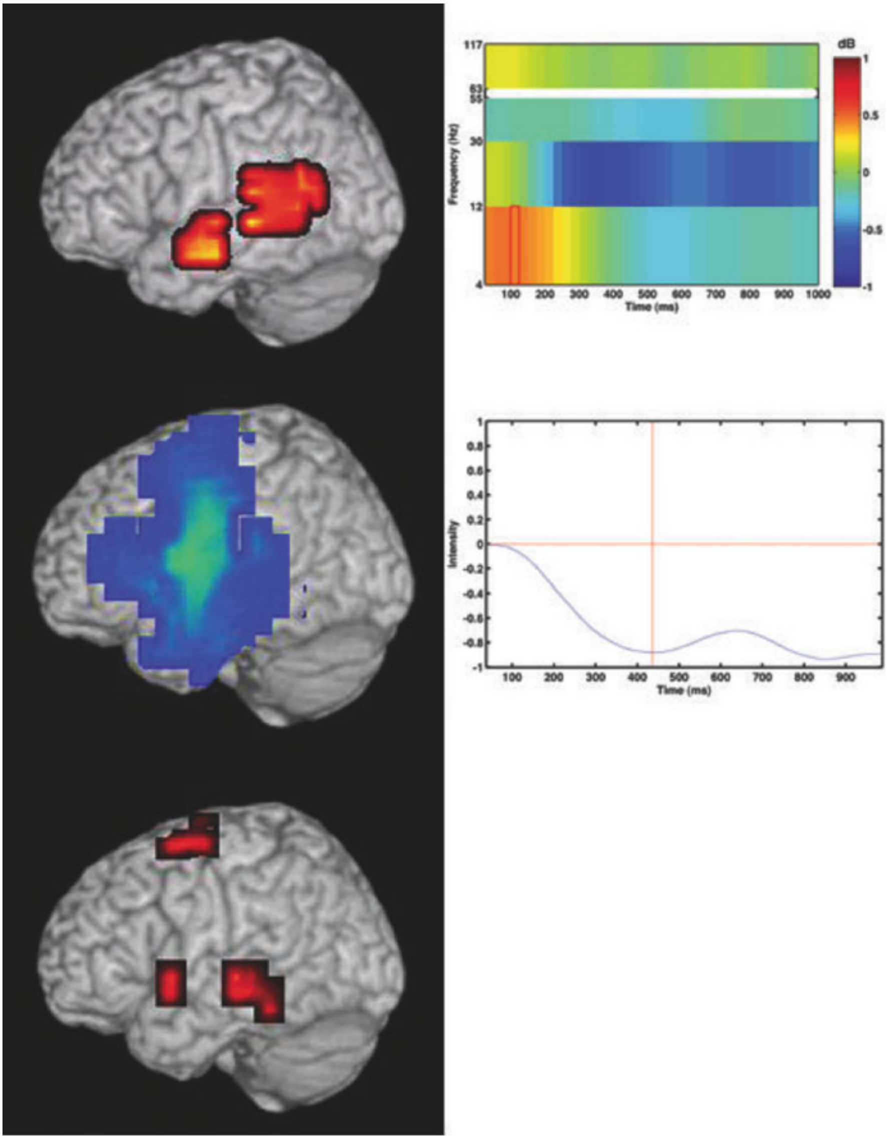


Fig. 5.4 The top panel is a full time-frequency induced oscillatory power modulation by speech syllables reconstructed using adaptive spatial filtering. The three-dimensional overlay rendering in the top row is evoked power in the theta/alpha band from 4–12 Hz, which corresponds to the evoked m100 response arising from activity in the superior temporal lobe. The middle row shows the beta-band desynchronization that follows, with the time course for superior temporal gyrus. Bottom row shows evoked gamma-band activity in the early time window showing synchronized motor activation. Time frequency response of auditory voxels are shown in top right, and time course in the middle shows beta-band power decreases following syllable onset

5.4.1 Transient Auditory Evoked Fields

Auditory evoked fields (AEFs) are cortical responses that are faithfully time locked to an externally presented auditory stimulus. There are a number of response components worth considering in the AEF that have been studied with MEG (see Fig. 5.1). The most dominant ones are the P1 (~50–100 ms), N100m (~100 ms), P2 (~100–200 ms), and N2 (~200 ms). Among these components, the N100m, also referred to as the M100 response component, is the most widely studied in MEG because of its ubiquity and by virtue of being the largest component with highest SNR. The M100 response arises from primary auditory cortex and its immediate environs on Heschl's gyrus (see Fig. 5.2) (Lutkenhoner et al., 2003).

5.4.2 Evoked Responses to Non-speech Acoustic Stimuli

Although a few studies have found a spatial tonotopic arrangement in auditory cortex using MEG, such organization has increased intra- and interindividual differences (Lutkenhoner et al., 2003), potentially accounting for why many studies have been unable to consistently report tonotopy using MEG. Although some studies have suggested an "amplitopic" organization in auditory cortex as demonstrated by MEG, the dipole location of the N100m shifts with changing stimulus intensity from 30 to 80 dB, whereby the depth of the N100m decreases with increasing sound intensity. However, this effect is potentially confounded by the estimation procedure. Increasing the amplitude of signals tends to result in deeper sources with dipole fitting procedures, and source depth is affected by SNR in the data (Pantev et al., 1989).

In contrast to spatial tonotopy, numerous studies demonstrate that changes in the carrier frequency of a tone affect the latency of the M100 response (Forss et al., 1993; Lutkenhoner et al., 2001; Krumbholz et al., 2003). Low-frequency tones have a longer latency when compared to higher frequency tones (Poeppel et al., 1996; Roberts & Poeppel, 1996). Further, increases in sound level for a simple tone at a fixed frequency cause the amplitude of the N100m to increase, while having no effect on latency; this growth in auditory evoked magnetic field amplitude is constant with low-frequency sounds (250–1000 Hz) and is less prominent with higher frequencies (>2000 Hz) (Soeta & Nakagawa, 2009).

In general, the pitch of acoustic stimuli is inversely related to the latency of the M100 response (Forss et al., 1993), implying that cortical elements at the source of the M100 response may be involved in pitch processing. To better isolate the pitch-onset response, Krumboltz et al. (2003) utilized regular-interval sounds that are able to introduce a pitch into the perception of sound without changing other components of the sound such as latency and amplitude and demonstrate that the latency and amplitude of the pitch-onset response varies with the pitch of the sound stimuli. Source localization with dipole fitting revealed that a pitch processing center lies approximately anterior and inferior to that of the N100m, possibly in the medial part of Heschl's gyrus (Krumbholz et al., 2003).

5.4.3 Effects of Stimulus Timing and Pattern on Early Response Components

The temporal pattern of sound stimuli influences auditory central processing (Carver et al., 2002; Rosburg et al., 2002). Following an initial sound stimulus, the auditory response to a second sound with the same spectral characteristics is recognizable with an interstimulus gap as short as 1 ms; and as that gap duration increases, so does the amplitude of the auditory evoked response field (Rupp et al., 2000). Further, we see a linear increase in the amplitude at the N100m as the interstimulus interval increases among a train of auditory tones presented at a specific rate (Carver et al., 2002). The mechanism of the effect may be due to the longer duration of time allowed to pass the refractory period that follows neuronal activation. The duration of the sound stimulus itself also has an effect on the AEMP. As stimulus duration increases, the N100m dipole location shifts more anterior and inferior, more in the right hemisphere than the left (Rosburg et al., 2002). Repetitive stimulation from a sound source leads to habituation of the N100m characterized by a decrease in amplitude and an inferior–superior dipole shift; however, interestingly, this habituation profile is not affected by acoustic stimulus duration (Rosburg et al., 2002). Although typical early response studies have focused on the transitions from silence to acoustic stimulation, some recent studies have examined transitions between different kinds of sounds. For instance, specific responses are observed for transitions between ordered versus disordered tone sequences, or from broadband noise stimuli to tones (Chait et al., 2007), or from a steady tone of one frequency to another (see Fig. 5.5).

5.4.4 Hemispheric Lateralization of Early Auditory Responses

Cerebral hemispheric lateralization refers to the asymmetric localization of cortical activity on either the right or left side of the brain. Classic theories indicate that speech is predominantly processed in the left hemisphere (Eulitz et al., 1995; Alho et al., 1998), whereas music is thought to be predominantly processed in the right (Zatorre et al., 1994; Griffiths et al., 1999; Zatorre et al., 2002). However, this hemispheric asymmetry may not be limited to the processing of highly complex sounds such as speech and music, but to the fundamental components of sound, particularly the spectral and temporal acoustic characteristics (Howard & Poeppel, 2009). Spectral changes in sound are principally processed in the right hemisphere whereas temporal changes are processed mainly in the left hemisphere (Okamoto et al., 2009). MEG reveals that frequency variations are also processed differently between the two hemispheres. In the right hemisphere, isofrequency bands for both 400- and 4000-Hz tones spatially migrate toward the anterolateral direction before the N100m peak, whereas in the left hemisphere, movements for 400-Hz and 4000-Hz tones are anterolateral and lateral, respectively (Ozaki et al., 2003). This difference perhaps reflects distinct functional roles in auditory information processing

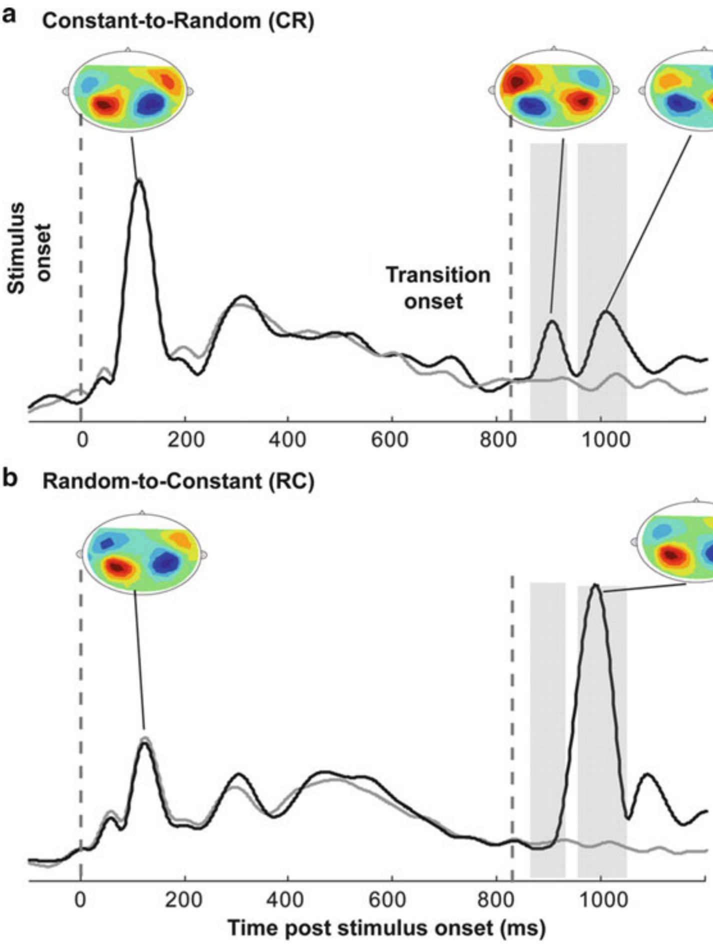


Fig. 5.5 (**a**) Global magnetic field power (root-mean-square of amplitude across all sensors) low-pass filtered at 10 Hz in response to a transition from a random stimulus to a constant stimulus. (**b**) Responses for transition from random stimuli to constant stimuli. The response for the no change control condition is plotted in gray. Contour maps at the critical time periods are also provided (7.5 fT/iso contour). Red, Source; blue, sink. Onset-response dynamics to CR and RC stimuli were comparable and characterized by a pronounced M100 onset response at ~110 ms after onset, with similar magnetic field distributions

between the two hemispheres. In normal healthy subjects, the M100 activation from a simple tone in the right hemisphere is located more anterior than in the left hemisphere. This expected auditory processing asymmetry was absent in autistic patients (Schmidt et al., 2009) and may be used to predict overall oral language ability in healthy and autistic patients (Oram Cardy et al., 2008).

Lateralization is also contingent on location of sound source relative to right versus left ear. When the N100m is measured from tones delivered alternately to the left and right ears, peak latencies are shorter in the contralateral hemisphere with a decrease more pronounced from stimuli presented to the left ear (Yu et al., 2007). Further, when comparing binaural to monoaural stimulation, there is symmetrical suppression of the right hemisphere under both conditions, whereas in the left hemisphere, responses are more suppressed for ipsilateral than for contralateral sounds (Fujiki et al., 2002). This suggests a right-ear dominance of left auditory cortex for nonspeech sounds. Other evidence also supports left hemisphere dominance during auditory processing in a noisy environment (Okamoto et al., 2007).

5.4.5 Mismatch Negativity Fields

The auditory mismatch negativity (MMN) is a component of the auditory evoked response to changes in stimulus patterns. MMN is generated by the brain's automatic response to a deviation from a standard stream of auditory stimuli corresponding to the behavioral discrimination threshold. To measure MMN, one must calculate the difference between the evoked responses elicited by standard versus deviant responses, specifically dynamics in latency, amplitude, and source localization. Studies of MMN have been a popular target in MEG research because it underlies cortical processes in central auditory processing and variations of auditory memory. The brain must maintain a certain length of auditory memory to recognize the presence of an "unusual" stimulus that otherwise does not fit the established memory trace (Näätänen et al., 2007).

A prerequisite for MMN generation is that before a deviant stimulus is presented, the central auditory system must first establish a memory stream of a "normal" auditory pattern. This standard auditory pattern may involve frequent stimuli with common characteristics in either interval time between tones, volume, duration, frequency, or intensity of sound, to name a few. After this auditory memory has been developed, deviant stimuli elicit this unique response. The MMN generally peaks at 150–250 ms following the onset of the deviant stimuli (Näätänen et al., 2007). Although some forms of MMN can be thought to be related to stimulus specific adaptation to the standards, not all MMN responses can be accounted for by this bottom-up perspective.

Dipole source localization of MMN indicates that changes in frequency, intensity, or duration are located at different areas. MMN dipoles of both frequency and duration deviants are inferior in location to intensity deviants, while MMN for duration and frequency differ in anterior–posterior direction (Rosburg, 2003). There also exists evidence that induced cortical responses are apparent in MMN paradigms. The source localization and changes in oscillatory activity elicited by the oddball task show that following deviant stimuli, changes in the delta frequency range occur in the frontocentral and parietal regions, theta and alpha range occurred over the dorsolateral and medial prefrontal cortex, and suppression in mu, beta, and low-gamma frequency range within bilateral central-Rolandic regions (Ishii et al., 2009).

5.4.6 Steady-State Evoked Responses

When continuous auditory stimuli are presented at fixed rate ranging from 20 to 40 Hz, they evoke what is called an auditory steady-state evoked response (aSSR), which is an end result of the superposition of multiple transient responses evoked by each presented stimuli. The 40-Hz auditory steady-state response was first demonstrated by EEG recordings from humans by Galambos et al. (1981) and several studies have examined the nature of this response using MEG. For example, children and adolescents with autism exhibit significantly reduced left-hemispheric steady-state gamma response power, whereas no difference was seen in the right hemisphere (Wilson et al., 2007). The aSSR dynamics for diotic and dichotic rhythmic stimulation show that sources are more anterior, inferior, and medial compared to the N100m sources. Further, there is a right hemispheric lateralization for the aSSR, which is consistent with the hypothesis that the right hemisphere is involved in processing of rhythmic sounds (Draganova et al., 2008).

aSSR is affected by changes in the sound localization. Interaural phase difference changes in the sound carrier from 0 to 180 degrees induces desynchronization of the ongoing aSSR, which is characterized by a decrement in aSSR amplitude 100 ms after changes in the interaural phase difference (Ross, 2008). In an elegant study of the aSSR, Gutschalk et al. studied the sustained field response to regular and irregular auditory click trains at three different sound intensities, and demonstrated that the sustained field within the lateral aspect of Heschl's gyrus was particularly sensitive to regularity of the click trains and not to sound level, whereas the region posterior to the first in planum temporale was more sensitive to sound level and not to regularity (Gutschalk et al., 2002). This provides evidence of two separate auditory sources responsible for processing pitch versus loudness. Apart from click trains, a great variety of other periodic stimuli have been used to elicit steady-state responses (Picton et al., 2003).Steady-state responses have recently been used profitably to test, for example, how simultaneous AM and FM modulation are encoded (Luo et al., 2006) and how long acoustic sequences (typical of speech or music) are reflected in the aSSR (Patel, 2003).

5.4.7 Evoked Responses to Speech Syllables

A range of MEG studies of evoked activity have revealed that phonological processing in auditory cortex may begin as early as 100 ms after sound onset, while semantic and lexical processing in auditory cortex starts between 200 and 300 ms after sound onset. Studies of syllables (ie /ba/, /da/, /ga/) have shown that the initial stages of speech-induced cortical activity manifest in MEG signals similar to those evoked by basic sounds, namely the bilateral N50m and N100m. In contrast to nonspeech sounds, the N100m elicited by speech sounds is often followed by a sustained activation, beginning at about 200 ms post-stimulus, peaking at 400 ms, and lasting for another 200–400 ms. This late, speech-specific activation is often referred to as the N400m (Helenius et al., 2002; Marinkovic et al., 2003; Biermann-Ruben et al., 2005). The shape, amplitude, and latency of the N100m may contain phonological information as well as context dependent semantic information (Shtyrov et al., 2010). In contrast, the later N400m has been shown to reflect semantic information as well as contextual information (Lau et al., 2008).

Observations have substantiated the assumption that the P50m and N100m response reflects an abstract phonological representational stage during speech perception (Ackermann et al., 2001; Tavabi et al., 2007). Earlier auditory evoked responses in the P50m waveform are sensitive to place-of-articulation features of vowels, thereby possibly serving as a segmentation and feature extractor to facilitate speech perception (Tavabi et al., 2007). The waveform of the P50m exhibits a skewed shape configuration characterized by an initial maximum peak followed by a knot over the contralateral (opposite cortex relative to monoaural sound stimulation) auditory cortex and a reversed pattern over the ipsilateral temporal lobe (Ackermann et al., 2001). Varying phonological information in the consonant and vowel component of the consonant–vowel syllable reveals a N100m source location difference along the anterior–posterior axis due to mutually exclusive places of articulation in the vowel of the syllable (Obleser et al., 2003). In addition, the N100m response is affected by whether the subject is passively or actively listening (subject asked to discriminate between syllables) (Poeppel et al., 1996). Under passive conditions, latencies and peaks of the N100m are bilaterally symmetrical, whereas in the active case, amplitude of the N100m in the left hemisphere is increased relative to the right hemisphere.

5.4.8 Oscillations Induced by Speech Syllables

The arrival of the speech sound signal in auditory cortex generates changes in neural oscillatory activity detectable by MEG that can be localized using time-frequency optimized adaptive spatial filtering methods as shown in Figure 5.4 (Dalal et al., 2008). In response to isolated speech syllables (/ba/ or /pa/ or /da/) starting around 50 ms after speech onset, the power in low-frequency (2–12 Hz) neural oscillations begins to increase in auditory cortex bilaterally, reflecting in part the m50 and m100 evoked activations. The spectral center of the activation appears somewhat variable between individuals. Concurrent with the low-frequency power increase, high gamma power over temporal lobes increases. The high gamma band power increases may reflect encoding of the phonological content of speech. Following the theta/alpha and high gamma band power increases, the beta-band undergoes a robust power decrease of auditory cortex, extending into medial temporal cortex and into inferior sensorimotor cortex (see Fig. 5.4). After the initial power increases, the power in the theta band tracks the temporal envelope of the perceived speech.

Past MEG studies have demonstrated a relationship between the intelligibility of speech and theta phase-speech envelope tracking. The theta phase pattern in auditory cortex distinguishes between different sentences, and is likely used by the brain to encode the syllabic structure of speech (Luo & Poeppel, 2007). Theta-phase speech envelope tracking persists when speech is compressed but still completely intelligible; however, it disappears when speech is compressed beyond complete intelligibility (Ahissar et al., 2001). Consistent with this idea, a recent study suggests that auditory cortical theta oscillations will track any sound with the same statistical acoustic properties as speech, and that theta tracking of such speech-like sound disappears when the sound is compressed beyond discriminability with other statistically similar sounds (Howard & Poeppel, 2010).

5.4.9 Responses to Vowels

The sound patterns encoding vowels consist of a fundamental frequency (F0) accompanied by harmonic components produced by resonance with the vocal cords. Similar to the way that the sonic characteristics of different musical instruments manifest in timbre, the vocal chords amplify specific harmonics of the voice F0, known as the formants (usually enumerated F1, F2, and F3), which carry information about the vowel being communicated. For example, the three English vowels /a/, /i/, and /u/ have the same fundamental frequency of about 100 Hz; however, each has different formant frequencies. The F1, F2, and F3 for the vowel /a/ are approximately 600 Hz, 1000 Hz, and 2500 Hz, respectively, and for /i/ are 200 Hz, 2300 Hz, and 3000 Hz, respectively. One widely utilized method to study speech-specific processing in the auditory cortex is to compare cortical activity in response to speech sounds, such as vowels, to complex nonspeech sounds that share the same dominant formant frequencies (F1, F2, F3) (Parviainen et al., 2005).

A left hemispheric lateralization commonly exists for both early and late evoked response components to vowels when compared to nonspeech sounds (Vihla & Salmelin, 2003; Parviainen et al., 2005). When comparing vowel acoustic stimuli to complex sounds with matching formant frequencies, the N100m response is much stronger for the former in the left hemisphere. The N100m response in the left hemisphere to simpler nonspeech sounds, defined as sharing only one formant frequency to the corresponding vowel, is even more weakened (Parviainen et al., 2005). This difference in N100m response between the vowel, complex sound, and simple sound was not evident in the right hemisphere. For vowels, the initial buildup of the N100m response was significantly steeper in the left than in right hemisphere, whereas this asymmetry was not apparent in nonspeech sound stimuli (Parviainen et al., 2005). Further, N100m amplitude is stronger in the left than in right hemisphere for vowels, compared to a lack of lateralization for simple tone stimuli (Gootjes et al., 1999; Tiitinen et al., 1999). N1m and the P2m latencies seem to peak earlier for tones than for vowels; however, there is no difference in source localization (Tiitinen et al., 1999).

The topographic arrangement of vowel representation in auditory cortex can be observed with MEG dipole fitting procedures. The distance between each vowel's N100m equivalent current dipole location within auditory cortex corresponds to the distances between corners of a vowel trapezium; however, the spatial configuration has high intersubject differences (Diesch et al., 1996). The latency, amplitude, and source locations of the N100m component can also differ among the vowels (Obleser et al., 2003; Obleser et al., 2006); specifically, the most dissimilar vowels are more spatially distant than more similar vowels. Topographic arrangement of speech sounds may be based on phonetic features only when intelligible (Cansino et al., 2003; Obleser et al., 2006). This topographic arrangement may be based on the F0 of these complex sounds (Cansino et al., 2003).

MEG has also helps understand how the fundamental and formant frequencies of vowels affect auditory speech representations. Speech sounds, with and without phonetics (same F0 with or without formant frequencies), share a constant N100m latency of approximately 120 ms while N100m latency for pure tones ranges from 120 to 160 ms, suggesting that the F0 contributes to the temporal dynamics of speech processing in auditory cortex (Makela et al., 2002). Interestingly, the source location and amplitude are not affected by the F0 of periodic vowels, but the latency increases as F0 frequency decreases, similar to nonspeech complex tones (Yrttiaho et al., 2008).

MEG accurately quantifies induced cortical responses to language stimuli. Eulitz et al. compared syllables to acoustically similar nonlanguage stimuli and revealed an enhancement of normalized spectral power at 240 ms latency in the 60- to 65-Hz band over the left hemisphere for the language condition and over the right hemisphere for the nonlanguage condition (Eulitz et al., 1996), but effects restricted to such narrow bands do not lend themselves to consistent interpretations about the underlying neural oscillations involved.

Functional coupling of alpha rhythms during dichotic listening of consonant–vowel syllables increases between left auditory cortex and Wernicke's area, while it decreases between the left and right auditory areas (Brancucci et al., 2008). Papanicolaou et al. compared the spatiotemporal dynamics of speech stimuli to analogous nonspeech stimuli and found that at a 60–130 ms latency, there is comparable primary auditory cortex activation bilaterally equal in strength in nonlanguage and language stimuli; however, at 130–800 ms, activation of the posterior portion of the superior temporal gyrus is greater in the left than in the right hemisphere for speech stimuli, while there is lack of this lateralization in nonspeech stimuli (Papanicolaou et al., 2003). This provides more evidence that activity in the supratemporal plane and cortex within the superior temporal sulcus, which are both components of the superior temporal gyrus, specialize in left hemispheric processing of speech sounds.

5.4.10 Mismatch Negativity to Speech Sounds

The MMN is also elicited when speech sounds are presented in a passive oddball task (Näätänen et al., 1997; Kraus et al., 1999; Näätänen & Alho, 2007). A vowel mismatch (/a/ against /e/) results in an enlarged N100m amplitude at the level of the supratemporal plane (Mathiak et al., 1999). Vowel deviants presented within a series of native-language vowel standards elicits larger MMN amplitude when it is a typical exemplar of the subject's native language versus when not in the native language (Näätänen & Alho, 1997; Sharma & Dorman, 1999; Szymanski et al., 1999). Infrequently presented vowels among repetitive vowels elicit a stronger MMN in the left compared to right auditory cortex with a source localization more posterior in the left than right (Koyama et al., 2000).

MEG is a useful tool for determining the effects on the MMN response to changes in formant frequencies. A change in F1 yields an enhancement of the N100m, which is linear to the spectral difference between the standard and deviant stimuli, whereas a change in F2 generates a nonlinear relationship (Mathiak et al., 2002). Vihla et al. studied whether spectral changes were interpreted differently in speech versus nonspeech sounds by comparing the MEG response for the MMN for natural vowels to sounds consisting of two pure tones that represent the lowest formant frequencies of the corresponding vowels. They demonstrated that the degree of spectral differences between standards and deviants are reflected by the MMF amplitude for nonspeech sounds and by the latency for vowel stimuli (Vihla et al., 2000). Makela et al. investigated the cortical dynamics underlying perception of gliding F0 for vowels and demonstrated that the N100m amplitude is highly sensitive to F0 variation. In contrast to vowel stimuli, corresponding tones has a more delayed N100m latency (Makela et al., 2004). Mismatch negativity with deviant vowel stimuli is significantly delayed in children with autism compared to normal healthy subjects (Oram Cardy et al., 2005).

5.4.11 Modulation of Auditory Cortical Responses During Speaking

Previous MEG studies have revealed a phenomenon called speaking-induced suppression: a reduced response in auditory cortex to self-produced speech, compared with its response to externally produced speech. These studies found a dampened auditory M100 response to a person's own voice when speaking compared to conditions in which a person listens to recorded speech being played back (Curio et al., 2000; Gunji et al., 2001; Houde et al., 2002). Auditory cortical activity was maximal when the recorded speech was different than the self-generated voices, but suppressed when both conditions were the same voice (Hirano et al., 1997). It is hypothesized that during speaking, actual incoming auditory stimulation is compared with a prediction derived from the efference copy of the motor output command, creating a feedback prediction error. It is this comparison that is hypothesized to be the principal cause of the speaking-induced suppression phenomenon seen in MEG studies (Fig. 5.6). Further, Ventura et al. sought to explore the effects of utterance rapidity and complexity on the speaking-induced suppression of M100 amplitudes. In this study, speaking-induced suppression (SIS) was greatest in simple, static utterances, but was significantly reduced as the utterances became more rapid and complex (Ventura et al., 2009). Presumably, the increased SIS was observable in the former condition because utterance was largely static and therefore easier to

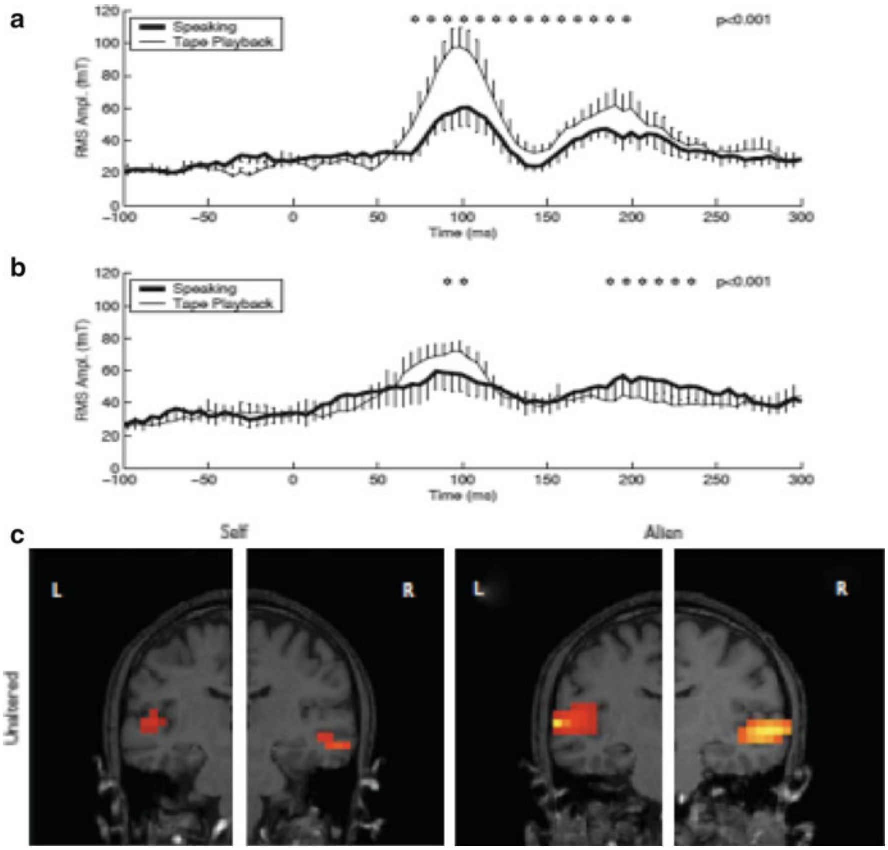


Fig. 5.6 (**a**) Speaking-induced suppression of responses in auditory cortex. The top row shows the RMS of left hemisphere sensors and the bottom row shows the right hemisphere data. Both rows show the response to self-produced speech (dark curve) when compared to listening to acoustically identical tape-recorded speech (light curve). Responses are significantly suppressed in both hemispheres, although larger in the left hemisphere. (**b**) Reconstruction of self-produced speech showing bilateral auditory cortex activation. (**c**) Reconstruction of tape-recorded speech also showing bilateral auditory cortical activation that is greater than during self-produced speech. Reconstructions shown for a single subject obtained by minimum-variance adaptive spatial filtering of data for each hemisphere separately, to circumvent sensitivity to correlated sources. (From Houde et al., 2002 and Heinks-Maldonado et al., 2007.)

produce and match, whereas in utterances that were more complex and rapid, auditory feedback predictions became more dynamic and more difficult to keep in temporal registry with the incoming auditory feedback. Further research in the speaking-induced suppression paradigm may benefit studies in disorders such as schizophrenia, whereby patients lack the ability to distinguish between internally and externally produced speech sounds (Heinks-Maldonado et al., 2007).

The idea of suppression of auditory cortical responses during speaking can be generalized to ideas of predictive coding, whereby predictable stimuli can suppress responses to arbitrary sounds in auditory cortex. The most striking general example of such a phenomenon is motor-induced suppression, defined by suppression of sensory responses to stimuli that are initially triggered by a self-initiated motor act. Martikainen et al. tested whether responses in human auditory cortex would differ to self-triggered (from button push) versus externally triggered tones (Martikainen et al., 2005). Sources of the M100 auditory response were significantly attenuated to self-triggered compared to externally triggered sounds. Aliu et al. examined the M100 auditory response to a simple tone triggered by a self-initiated button press (Aliu et al., 2009) and found that such suppression develops over about 100 trials, can also develop if a delay is introduced between the motor act and the sensory stimulus, and that suppression occurs in excess of auditory habituation.

Taken together with MMN studies, it can be stated that auditory cortical responses attenuate for predictable stimuli and are geared toward detection of novel stimulus patterns.

5.5 Auditory Cortical Plasticity Assessed by MEG

The term auditory plasticity encompasses the ability of the brain to adapt and learn from changes in the acoustic environment. Such dynamic changes can occur rapidly (Pantev et al., 1999; Ross & Tremblay, 2009; Sanders et al., 2009) or gradually over a longer term (Shahin et al., 2005; Pantev et al., 2009; Trainor et al., 2009). MEG has provided effective ways to investigate the plastic changes that occur in auditory cortex.

MEG studies demonstrate the underlying neural mechanisms of rapid plasticity occurring in auditory tasks (Hairston & Nagarajan, 2007). The N100m amplitude decreases continuously as a simple tone is repeatedly played; however, it recovers to baseline amplitude once a new session is restarted. In contrast, the P200m amplitude seems to increase after multiple sessions of repeated acoustic stimuli, suggesting that the long-lasting increase in the P200m compared to the N100m may be a neural correlate of perceptual learning and training. This P200m enhancement decreased with age (Ross & Tremblay, 2009).

When discriminating between acoustic stimuli varying in frequency or intensity, cortical plasticity is evident with improved discrimination characterized by changes in the amplitude of the M100 response in primary auditory cortex (Cansino et al., 1997). Rapid changes can occur in the tuning of neurons when specific frequencies in the acoustic environment are removed as well (Pantev et al., 1999). Presentation order of acoustic input also influences performance accuracy in a discrimination task (Menning et al., 2002; Hairston & Nagarajan, 2007). Specifically, the M300 component of the AEF may be a neural correlate of plasticity for this time-order error phenomenon (Hairston & Nagarajan, 2007).

Auditory cortex plasticity is also evident when listeners are trained to recognize a sequence of nonsense sounds that are otherwise not easily rehearsed (Sanders et al., 2009). In this case, approximately 40 ms after acoustic onset, only the initial sound of the sequence elicited plastic changes characterized by an enhanced amplitude in its auditory evoked field. Learning foreign language speech sounds also induces plastic changes in the brain (Menning et al., 2002). In Japanese, mora is a temporal unit that divides words into isochronous segments. Learning to discriminate these mora is associated with an increase in amplitude and decrease in latency in the mismatch field.

Plastic changes in the auditory system are also evident in long-term learning processes such as musical training. When comparing cortical response to musical sounds in musicians versus nonmusicians, the P200m amplitude, but not the N100m, was generally enhanced and increased with spectral complexity in the musicians (Shahin et al., 2005, 2007). Perhaps, the P200m corresponds to features of acoustic stimuli experience in musical practice while the N100m waveform is unaffected. Interestingly, P200m enhancement also varies according to musical instrument type in children (Shahin et al., 2004). Timbre-specific enhancement is also seen in musicians, as demonstrated by MEG (Pantev et al., 2001). Further, musicians show larger MMNm responses to changes in melodic contour than nonmusicians do. Both groups, however, show similar changes in the frequency of a pure tone (Pantev et al., 2003).

Pitch discrimination is a skill more fine tuned in musicians than in nonmusicians. The P200m and N100m are sensitive to neural remodeling from pitch discrimination in nonmusicians. When comparing the plastic changes that occur with pitch discrimination training, musicians have larger N100m and P100m amplitudes (Shahin et al., 2003). Absolute pitch is a term that describes the ability of a person to recreate a musical note without the help of an external reference, a skill trained musicians can develop. Hirata et al. compared MEG response of musicians with absolute pitch to nonmusicians while they received auditory stimuli (Hirata et al., 1999). The equivalent current dipole (ECD) for the noise burst showed significant different in spatial location between both groups, which may be a result of cortical plasticity produced by musical training. Musicians exhibit enlarged cortical representation of musical tones compared to pure tones, which correlates with the age at which the musician began to practice (Pantev et al., 2003).

Musical training affects the oscillatory networks in the brain. Phase-locking in the beta (15–30 Hz) and high-gamma (30–70 Hz) range matures later and was stronger in music sound frequencies when compared to pure tones matched in fundamental frequency. Further, phase-locking strengthened with age (Shahin et al., 2010). Induced gamma band response to musical sounds is enhanced in musicians when compared to nonmusicians, and this plastic change develops in children after 1 year of musical training (Trainor et al., 2009).

Performing and listening to music is a complex process, which requires multimodal information processing not just in auditory cortex, but in the somatosensory networks as well. Therefore, it is no surprise that cortical plasticity to musical training is exhibited not just in auditory regions but also in other cortical areas. The somatosensory cortex may induce cross-modal plasticity and affect changes in the primary sensory cortices; therefore, training-induced musicians may have qualitative differences in the way they process multisensory information. For example, when lips of trumpet players are stimulated at the same time as a trumpet tone, activation in the somatosensory cortex is increased more than during the sum of separate lip and trumpet tone stimulation (Pantev et al., 2003).

The brain also has the capacity to undergo plasticity after a postcerebral insult (Breier et al., 2006; Rossini et al., 2007). This provides evidence that reorganization of auditory cortex and association cortices can occur even in adults. Breier et al. studied patients with epilepsy, surgical resection, and stroke and looked at plasticity associated with a recognition task for spoken words (Breier et al., 2006). Here they found that increased activation in the right hemisphere postoperatively was associated with greater relative activation preoperatively. Patients with epilepsy secondary to a neoplasm or to mesial temporal sclerosis exhibited differences in shifts of their language function. In addition, patients with stroke-induced aphasia had a more bilateral and diffuse overall activation profile with the language-dominant hemisphere.

5.6 Summary and Conclusions

In this chapter, the technological capabilities and limits of auditory studies with MEG are discussed. Although the majority of auditory studies using MEG have been restricted either to sensor analyses or dipole fitting procedures, much has been learned about auditory cortical representations using MEG. With the advent of more advanced and sophisticated techniques for reconstructing auditory cortical responses with greater fidelity and robustness, it is expected that the next wave of auditory studies using MEG will exploit the full power of reconstruction algorithms for MEG imaging and pave the way for greater understanding of auditory cortical activity.

References

Ackermann, H., Hertrich, I., Mathiak, K., & Lutzenberger, W. (2001). Contralaterality of cortical auditory processing at the level of the M50/M100 complex and the mismatch field: A whole-head magnetoencephalography study. *NeuroReport*, 12(8), 1683–1687.

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences*, *of the USA*, 98(23), 13367–13372.

Alho, K., Connolly, J. F., Cheour, M., Lehtokoski, A., Huotilainen, M., Virtanen, J., et al. (1998). Hemispheric lateralization in preattentive processing of speech sounds. *Neuroscience Letters*, 258(1), 9–12.

Aliu, S. O., Houde, J. F., & Nagarajan, S. S. (2009). Motor-induced suppression of the auditory cortex. *Journal of Cognitive Neuroscience*, 21(4), 791–802.

Biermann-Ruben, K., Salmelin, R., & Schnitzler, A. (2005). Right rolandic activation during speech perception in stutterers: A MEG study. *NeuroImage*, 25(3), 793–801.

Brancucci, A., Penna, S. D., Babiloni, C., Vecchio, F., Capotosto, P., Rossi, D., et al. (2008). Neuromagnetic functional coupling during dichotic listening of speech sounds. *Human Brain Mapping*, 29(3), 253–264.

Breier, J. I., Billingsley-Marshall, R., Pataraia, E., Castillo, E. M., & Papanicolaou, A. C. (2006). Magnetoencephalographic studies of language reorganization after cerebral insult. *Archives in Physical Medicine and Rehabilitation*, 87(12 Supplement 2), S77–83.

Cansino, S., & Williamson, S. J. (1997). Neuromagnetic fields reveal cortical plasticity when learning an auditory discrimination task. *Brain Research*, 764(1–2), 53–66.

Cansino, S., Ducorps, A., & Ragot, R. (2003). Tonotopic cortical representation of periodic complex sounds. *Human Brain Mapping*, 20(2), 71–81.

Carver, F. W., Fuchs, A., Jantzen, K. J., & Kelso, J. A. (2002). Spatiotemporal analysis of the neuromagnetic response to rhythmic auditory stimulation: Rate dependence and transient to steady-state transition. *Clinical Neurophysiology*, 113(12), 1921–1931.

Chait, M., Poeppel, D., de Cheveigne, A., & Simon, J. Z. (2007). Processing asymmetry of transitions between order and disorder in human auditory cortex. *Journal of Neuroscience*, 27(19), 5207–5214.

Curio, G., Neuloh, G., Numminen, J., Jousmaki, V., & Hari, R. (2000). Speaking modifies voice-evoked activity in the human auditory cortex. *Human Brain Mapping*, 9(4), 183–191.

Dalal, S. S., Sekihara, K., & Nagarajan, S. S. (2006). Modified beamformers for coherent source region suppression. *IEEE Transactions in Biomedical Engineering*, 53(7), 1357–1363.

Dalal, S. S., Guggisberg, A. G., Edwards, E., Sekihara, K., Findlay, A. M., Canolty, R. T., et al. (2008). Five-dimensional neuroimaging: Localization of the time-frequency dynamics of cortical activity. *NeuroImage*, 40(4), 1686–1700.

Darvas, F., Pantazis, D., Kucukaltun-Yildirim, E., & Leahy, R. M. (2004). Mapping human brain function with MEG and EEG: Methods and validation. *NeuroImage*, 23(Supplemen*t* 1), S289–299.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9–21.

Diesch, E., Eulitz, C., Hampson, S., & Ross, B. (1996). The neurotopography of vowels as mirrored by evoked magnetic field measurements. *Brain and Language*, 53(2), 143–168.

Draganova, R., Ross, B., Wollbrink, A., & Pantev, C. (2008). Cortical steady-state responses to central and peripheral auditory beats. *Cerebral Cortex*, 18(5), 1193–1200.

Eulitz, C., Diesch, E., Pantev, C., Hampson, S., & Elbert, T. (1995). Magnetic and electric brain activity evoked by the processing of tone and vowel stimuli. *Journal of Neuroscience*, 15(4), 2748–2755.

Eulitz, C., Maess, B., Pantev, C., Friederici, A. D., Feige, B., & Elbert, T. (1996). Oscillatory neuromagnetic activity induced by language and non-language stimuli. *Brain Research: Cognitive Brain Research*, 4(2), 121–132.

Forss, N., Makela, J. P., McEvoy, L., & Hari, R. (1993). Temporal integration and oscillatory responses of the human auditory cortex revealed by evoked magnetic fields to click trains. *Hearing Research*, 68(1), 89–96.

Fujiki, N., Jousmaki, V., & Hari, R. (2002). Neuromagnetic responses to frequency-tagged sounds: A new method to follow inputs from each ear to the human auditory cortex during binaural hearing. *Journal of Neuroscience*, 22(3), RC205.

Galambos, R., Makeig, S., & Talmachoff, P. J. (1981). A 40-Hz auditory potential recorded from the human scalp. *Proceedings of the National Academy of Sciences*, *of the USA*, 78(4), 2643–2647.

Gootjes, L., Raij, T., Salmelin, R., & Hari, R. (1999). Left-hemisphere dominance for processing of vowels: A whole-scalp neuromagnetic study. *NeuroReport*, 10(14), 2987–2991.

Griffiths, T. D., Johnsrude, I., Dean, J. L., & Green, G. G. (1999). A common neural substrate for the analysis of pitch and duration pattern in segmented sound? *NeuroReport*, 10(18), 3825–3830.

Gunji, A., Hoshiyama, M., & Kakigi, R. (2001). Auditory response following vocalization: A magnetoencephalographic study. *Clinical Neurophysiology*, 112(3), 514–520.

Gutschalk, A., Patterson, R. D., Rupp, A., Uppenkamp, S., & Scherg, M. (2002). Sustained magnetic fields reveal separate sites for sound level and temporal regularity in human auditory cortex. *NeuroImage*, 15(1), 207–216.

Hairston, I. S., & Nagarajan, S. S. (2007). Neural mechanisms of the time-order error: An MEG study. *Journal of Cognitive Neuroscience*, 19(7), 1163–1174.

Heinks-Maldonado, T. H., Mathalon, D. H., Houde, J. F., Gray, M., Faustman, W. O., & Ford, J. M. (2007). Relationship of imprecise corollary discharge in schizophrenia to auditory hallucinations. *Archives in General Psychiatry*, 64(3), 286–296.

Helenius, P., Salmelin, R., Service, E., Connolly, J. F., Leinonen, S., & Lyytinen, H. (2002). Cortical activation during spoken-word segmentation in nonreading-impaired and dyslexic adults. *Journal of Neuroscience*, 22(7), 2936–2944.

Hillebrand, A., & Barnes, G. R. (2002). A quantitative assessment of the sensitivity of whole-head MEG to activity in the adult human cortex. *NeuroImage*, 16(3 Pt 1), 638–650.

Hirano, S., Kojima, H., Naito, Y., Honjo, I., Kamoto, Y., Okazawa, H., et al. (1997). Cortical processing mechanism for vocalization with auditory verbal feedback. *NeuroReport*, 8(9–10), 2379–2382.

Hirata, Y., Kuriki, S., & Pantev, C. (1999). Musicians with absolute pitch show distinct neural activities in the auditory cortex. *NeuroReport*, 10(5), 999–1002.

Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: An MEG study. *Journal of Cognitive Neuroscience*, 14(8), 1125–1138.

Howard, M. F., & Poeppel, D. (2009). Hemispheric asymmetry in mid and long latency neuromagnetic responses to single clicks. *Hearing Research*, 257(1–2), 41–52.

Howard, M. F., & Poeppel, D. (2010). Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *Journal of Neurophysiology*, 104(5), 2500–2511.

Ishii, R., Canuet, L., Herdman, A., Gunji, A., Iwase, M., Takahashi, H., et al. (2009). Cortical oscillatory power changes during auditory oddball task revealed by spatially filtered magnetoencephalography. *Clinical Neurophysiology*, 120(3), 497–504.

Koyama, S., Gunji, A., Yabe, H., Oiwa, S., Akahane-Yamada, R., Kakigi, R., & Näätänen, R. (2000). Hemispheric lateralization in an analysis of speech sounds. Left hemisphere dominance replicated in Japanese subjects. *Brain Researh: Cognitive Brain Research*, 10(1–2), 119–124.

Kraus, N., Koch, D. B., McGee, T. J., Nicol, T. G., & Cunningham, J. (1999). Speech-sound discrimination in school-age children: Psychophysical and neurophysiologic measures. *Journal of Speech Language and Hearing Research*, 42(5), 1042–1060.

Krumbholz, K., Patterson, R. D., Seither-Preisler, A., Lammertmann, C., & Lutkenhoner, B. (2003). Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. *Cerebral Cortex*, 13(7), 765–772.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience*, 9(12), 920–933.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412(6843), 150–157.

Luo, H., Wang, Y., Poeppel, D., & Simon, J. Z. (2006). Concurrent encoding of frequency and amplitude modulation in human auditory cortex: MEG evidence. *Journal of Neurophysiology*, 96(5), 2712–2723.

Luo, H., Wang, Y., Poeppel, D., & Simon, J. Z. (2007). Concurrent encoding of frequency and amplitude modulation in human auditory cortex: Encoding transition. *Journal of Neurophysiology*, 98(6), 3473–3485.

Lutkenhoner, B., Lammertmann, C., & Knecht, S. (2001). Latency of auditory evoked field deflection N100m ruled by pitch or spectrum? *Audiology and Neurootology*, 6(5), 263–278.

Lutkenhoner, B., Krumbholz, K., Lammertmann, C., Seither-Preisler, A., Steinstrater, O., & Patterson, R. D. (2003). Localization of primary auditory cortex in humans by magnetoencephalography. *NeuroImage*, 18(1), 58–66.

Makeig, S., Jung, T. P., Bell, A. J., Ghahremani, D., & Sejnowski, T. J. (1997). Blind separation of auditory event-related brain responses into independent components. *Proceedings of the National Academy of Sciences of the USA*, 94(20), 10979–10984.

Makela, A. M., Alku, P., Makinen, V., Valtonen, J., May, P., & Tiitinen, H. (2002). Human cortical dynamics determined by speech fundamental frequency. *NeuroImage*, 17(3), 1300–1305.

Makela, A. M., Alku, P., Makinen, V., & Tiitinen, H. (2004). Glides in speech fundamental frequency are reflected in the auditory N1m response. *NeuroReport*, 15(7), 1205–1208.

Marinkovic, K., Dhond, R. P., Dale, A. M., Glessner, M., Carr, V., & Halgren, E. (2003). Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron*, 38(3), 487–497.

Martikainen, M. H., Kaneko, K., & Hari, R. (2005). Suppressed responses to self-triggered sounds in the human auditory cortex. *Cerebral Cortex*, 15(3), 299–302.

Mathiak, K., Hertrich, I., Lutzenberger, W., & Ackermann, H. (1999). Preattentive processing of consonant vowel syllables at the level of the supratemporal plane: A whole-head magnetencephalography study. *Brain Research: Cognitve Brain Research*, 8(3), 251–257.

Mathiak, K., Hertrich, I., Lutzenberger, W., & Ackermann, H. (2002). The influence of critical bands on neuromagnetic fields evoked by speech stimuli in humans. *Neuroscience Letters*, 329(1), 29–32.

Menning, H., Imaizumi, S., Zwitserlood, P., & Pantev, C. (2002). Plasticity of the human auditory cortex induced by discrimination learning of non-native, mora-timed contrasts of the Japanese language. *Learning and Memory*, 9(5), 253–267.

Mosher, J. C., Leahy, R. M., & Lewis, P. S. (1999a). EEG and MEG: Forward solutions for inverse methods. *IEEE Transactions in Biomedical Engineering*, 46(3), 245–259.

Mosher, J. C., Baillet, S., & Leahy, R. M. (1999b). EEG source localization and imaging using multiple signal classification approaches. *Clinical Neurophysiology*, 16(3), 225–238.

Näätänen, R., & Alho, K. (1997). Mismatch negativity—the measure for central sound representation accuracy. *Audiology and Neurootology*, 2(5), 341–353.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, 118(12), 2544–2590.

Nagarajan, S. S., Attias, H. T., Hild, K. E., 2nd, & Sekihara, K. (2006). A graphical model for estimating stimulus-evoked brain responses from magnetoencephalography data with large background brain activity. *NeuroImage*, 30(2), 400–416.

Nagarajan, S. S., Attias, H. T., Hild, K. E., 2nd, & Sekihara, K. (2007). A probabilistic algorithm for robust interference suppression in bioelectromagnetic sensor data. *Statistics in Medicine*, 26(21), 3886–3910.

Niessing, J., Ebisch, B., Schmidt, K. E., Niessing, M., Singer, W., & Galuske, R. A. (2005). Hemodynamic signals correlate tightly with synchronized gamma oscillations. *Science*, 309(5736), 948–951.

Nunez, P. L., & Srinivasan, R. (2006). A theoretical basis for standing and traveling brain waves measured with human EEG with implications for an integrated consciousness. *Clinical Neurophysiology*, 117(11), 2424–2435.

Obleser, J., Lahiri, A., & Eulitz, C. (2003). Auditory-evoked magnetic field codes place of articulation in timing and topography around 100 milliseconds post syllable onset. *NeuroImage*, 20(3), 1839–1847.

Obleser, J., Scott, S. K., & Eulitz, C. (2006). Now you hear it, now you don't: Transient traces of consonants and their nonspeech analogues in the human brain. *Cerebral Cortex*, 16(8), 1069–1076.

Okamoto, H., Stracke, H., Ross, B., Kakigi, R., & Pantev, C. (2007). Left hemispheric dominance during auditory processing in a noisy environment. *Biomed Central Biology*, 5, 52.

Okamoto, H., Stracke, H., Draganova, R., & Pantev, C. (2009). Hemispheric asymmetry of auditory evoked fields elicited by spectral versus temporal stimulus change. *Cerebral Cortex*, 19(10), 2290–2297.

Oram Cardy, J. E., Flagg, E. J., Roberts, W., & Roberts, T. P. (2005). Delayed mismatch field for speech and non-speech sounds in children with autism. *NeuroReport*, 16(5), 521–525.

Oram Cardy, J. E., Flagg, E. J., Roberts, W., & Roberts, T. P. (2008). Auditory evoked fields predict language ability and impairment in children. *International Journal of Psychophysiology*, 68(2), 170–175.

Ozaki, I., Suzuki, Y., Jin, C. Y., Baba, M., Matsunaga, M., & Hashimoto, I. (2003). Dynamic movement of N100m dipoles in evoked magnetic field reflects sequential activation of isofrequency bands in human auditory cortex. *Clinical Neurophysiology*, 114(9), 1681–1688.

Palva, S., Palva, J. M., Shtyrov, Y., Kujala, T., Ilmoniemi, R. J., Kaila, K., & Näätänen, R. (2002). Distinct gamma-band evoked responses to speech and non-speech sounds in humans. *Journal of Neuroscience*, 22(4), RC211.

Pantev, C., Hoke, M., Lehnertz, K., & Lutkenhoner, B. (1989). Neuromagnetic evidence of an amplitopic organization of the human auditory cortex. *Electroencephalography and Clinical Neurophysiology*, 72(3), 225–231.

Pantev, C., Wollbrink, A., Roberts, L. E., Engelien, A., & Lutkenhoner, B. (1999). Short-term plasticity of the human auditory cortex. *Brain Research*, 842(1), 192–199.

Pantev, C., Roberts, L. E., Schulz, M., Engelien, A., & Ross, B. (2001). Timbre-specific enhancement of auditory cortical representations in musicians. *NeuroReport*, 12(1), 169–174.

Pantev, C., Ross, B., Fujioka, T., Trainor, L. J., Schulte, M., & Schulz, M. (2003). Music and learning-induced cortical plasticity. *Annals of New York Academy of Sciences*, 999, 438–450.

Pantev, C., Lappe, C., Herholz, S. C., & Trainor, L. (2009). Auditory-somatosensory integration and cortical plasticity in musical training. *Annals of the New York Academy of Sciences*, 1169, 143–150.

Papanicolaou, A. C., Castillo, E., Breier, J. I., Davis, R. N., Simos, P. G., & Diehl, R. L. (2003). Differential brain activation patterns during perception of voice and tone onset time series: A MEG study. *NeuroImage*, 18(2), 448–459.

Parviainen, T., Helenius, P., & Salmelin, R. (2005). Cortical differentiation of speech and non-speech sounds at 100 ms: Implications for dyslexia. *Cerebral Cortex*, 15(7), 1054–1063.

Patel, A. D. (2003). Rhythm in language and music: Parallels and differences. *Annals of the New York Academy of Sciences*, 999, 140–143.

Patel, A. D., & Balaban, E. (2004). Human auditory cortical dynamics during perception of long acoustic sequences: Phase tracking of carrier frequency by the auditory steady-state response. *Cerebral Cortex*, 14(1), 35–46.

Picton, T. W., John, M. S., Purcell, D. W., & Plourde, G. (2003). Human auditory steady-state responses: The effects of recording technique and state of arousal. *Anesthesia and Analgesia*, 97(5), 1396–1402.

Poeppel, D., Yellin, E., Phillips, C., Roberts, T. P., Rowley, H. A., Wexler, K., & Marantz, A. (1996). Task-induced asymmetry of the auditory evoked M100 neuromagnetic field elicited by speech sounds. *Brain Research: Cognitve Brain Research*, 4(4), 231–242.

Quraan, M. A., & Cheyne, D. (2010). Reconstruction of correlated brain activity with adaptive spatial filters in MEG. *NeuroImage*, 49(3), 2387–2400.

Roberts, T. P., & Poeppel, D. (1996). Latency of auditory evoked M100 as a function of tone frequency. *NeuroReport*, 7(6), 1138–1140.

Rosburg, T. (2003). Left hemispheric dipole locations of the neuromagnetic mismatch negativity to frequency, intensity and duration deviants. *Brain Research: Cognitive Brain Research*, 16(1), 83–90.

Rosburg, T., Haueisen, J., & Sauer, H. (2002). Stimulus duration influences the dipole location shift within the auditory evoked field component N100m. *Brain Topography*, 15(1), 37–41.

Ross, B., & Tremblay, K. (2009). Stimulus experience modifies auditory neuromagnetic responses in young and older listeners. *Hearing Research*, 248(1–2), 48–59.

Ross, B., Tremblay, K. L., & Picton, T. W. (2007). Physiological detection of interaural phase differences. *Journal of the Acoustical Society of America*, 121(2), 1017–1027.

Rossini, P. M., Altamura, C., Ferreri, F., Melgari, J. M., Tecchio, F., Tombini, M., et al. (2007). Neuroimaging experimental studies on brain plasticity in recovery from stroke. *Europa Medicophysica*, 43(2), 241–254.

Rupp, A., Hack, S., Gutschalk, A., Schneider, P., Picton, T. W., Stippich, C., & Scherg, M. (2000). Fast temporal interactions in human auditory cortex. *NeuroReport*, 11(17), 3731–3736.

Sanders, L. D., Ameral, V., & Sayles, K. (2009). Event-related potentials index segmentation of nonsense sounds. *Neuropsychologia*, 47(4), 1183–1186.

Schmidt, G. L., Rey, M. M., Oram Cardy, J. E., & Roberts, T. P. (2009). Absence of M100 source asymmetry in autism associated with language functioning. *NeuroReport*, 20(11), 1037–1041.

Sekihara, K., & Nagarajan, S. S. (2008). *Adaptive spatial filters for electromagnetic brain imaging*. New York: Springer.

Shahin, A., Bosnyak, D. J., Trainor, L. J., & Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *Journal of Neuroscience*, 23(13), 5545–5552.

Shahin, A., Roberts, L. E., & Trainor, L. J. (2004). Enhancement of auditory cortical development by musical experience in children. *NeuroReport*, 15(12), 1917–1921.

Shahin, A. J., Roberts, L. E., Pantev, C., Trainor, L. J., & Ross, B. (2005). Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. *NeuroReport*, 16(16), 1781–1785.

Shahin, A. J., Roberts, L. E., Miller, L. M., McDonald, K. L., & Alain, C. (2007). Sensitivity of EEG and MEG to the N1 and P2 auditory evoked responses modulated by spectral complexity of sounds. *Brain Topography*, 20(2), 55–61.

Shahin, A. J., Trainor, L. J., Roberts, L. E., Backer, K. C., & Miller, L. M. (2010). Development of auditory phase-locked activity for music sounds. *Journal of Neurophysiology*, 103(1), 218–229.

Sharma, A., & Dorman, M. F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *Journal of the Acoustical Soceity of America*, 106(2), 1078–1083.

Shtyrov, Y., Kujala, T., & Pulvermuller, F. (2010). Interactions between language and attention systems: Early automatic lexical processing? *Journal of Cognitive Neuroscience*, 22(7), 1465–1478.

Singh, K. D., Barnes, G. R., Hillebrand, A., Forde, E. M., & Williams, A. L. (2002). Task-related changes in cortical synchronization are spatially coincident with the hemodynamic response. *NeuroImage*, 16(1), 103–114.

Singh, K. D., Barnes, G. R., & Hillebrand, A. (2003). Group imaging of task-related changes in cortical synchronisation using nonparametric permutation testing. *NeuroImage*, 19(4), 1589–1601.

Soeta, Y., & Nakagawa, S. (2009). Sound level-dependent growth of N1m amplitude with low and high-frequency tones. *NeuroReport*, 20(6), 548–552.

Szymanski, M. D., Yund, E. W., & Woods, D. L. (1999). Phonemes, intensity and attention: Differential effects on the mismatch negativity (MMN). *Journal of the Acoustical Society of America*, 106(6), 3492–3505.

Tavabi, K., Obleser, J., Dobel, C., & Pantev, C. (2007). Auditory evoked fields differentially encode speech features: An MEG investigation of the P50m and N100m time courses during syllable processing. *European Journal of Neuroscience*, 25(10), 3155–3162.

Tiitinen, H., Sivonen, P., Alku, P., Virtanen, J., & Näätänen, R. (1999). Electromagnetic recordings reveal latency differences in speech and tone processing in humans. *Brain Research: Cognitive Brain Research*, 8(3), 355–363.

Trainor, L. J., Shahin, A. J., & Roberts, L. E. (2009). Understanding the benefits of musical training: Effects on oscillatory brain activity. *Annals of the New York Academy of Sciences*, 1169, 133–142.

Ventura, M. I., Nagarajan, S. S., & Houde, J. F. (2009). Speech target modulates speaking induced suppression in auditory cortex. *BioMed Central Neuroscience*, 10, 58.

Vihla, M., & Salmelin, R. (2003). Hemispheric balance in processing attended and non-attended vowels and complex tones. *Brain Research: Cognitve Brain Research*, 16(2), 167–173.

Vihla, M., Lounasmaa, O. V., & Salmelin, R. (2000). Cortical processing of change detection: Dissociation between natural vowels and two-frequency complex tones. *Proceedings of the National Academy of Sciences of the USA*, 97(19), 10590–10594.

Vrba, J., & Robinson, S. E. (2002). SQUID sensor array configurations for magnetoencephalography applications. *Superconducting Science and Technology*, 15(9), 51–89.

Wilson, T. W., Rojas, D. C., Reite, M. L., Teale, P. D., & Rogers, S. J. (2007). Children and adolescents with autism exhibit reduced MEG steady-state gamma responses. *Biological Psychiatry*, 62(3), 192–197.

Wipf, D., & Nagarajan, S. (2008). A unified Bayesian framework for MEG/EEG source imaging. *NeuroImage*, 44(3):947–66..

Wipf, D. P., Owen, J. P., Attias, H. T., Sekihara, K., & Nagarajan, S. S. (2010). Robust Bayesian estimation of the location, orientation, and time course of multiple correlated neural sources using MEG. *NeuroImage*, 49(1), 641–655.

Yrttiaho, S., Tiitinen, H., May, P. J., Leino, S., & Alku, P. (2008). Cortical sensitivity to periodicity of speech sounds. *Journal of the Acoustical Soceity of America*, 123(4), 2191–2199.

Yu, H. Y., Chen, J. T., Wu, Z. A., Yeh, T. C., Ho, L. T., & Lin, Y. Y. (2007). Side of the stimulated ear influences the hemispheric balance in coding tonal stimuli. *Neurological Research*, 29(5), 517–522.

Zatorre, R. J., Evans, A. C., & Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *Journal of Neuroscience*, 14(4), 1908–1919.

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Science*, 6(1), 37–46.

Zumer, J. M., Attias, H. T., Sekihara, K., & Nagarajan, S. S. (2007). A probabilistic algorithm integrating source localization and noise suppression for MEG and EEG data. *NeuroImage*, 37(1),102–115.

Zumer, J. M., Attias, H. T., Sekihara, K., & Nagarajan, S. S. (2008). Probabilistic algorithms for MEG/EEG source reconstruction using temporal basis functions learned from data. *NeuroImage*, 41(3), 924–940.

Chapter 6
Hemodynamic Imaging: Functional Magnetic Resonance Imaging

Thomas M. Talavage, Ingrid S. Johnsrude, and Javier Gonzalez-Castillo

6.1 Introduction

In the last few decades, structural and functional imaging methods have revolutionized cognitive neuroscience, providing a window into the functional organization and behavior of the human brain (review: Raichle, 2009). Neuroscientists no longer need rely on postmortem analysis of cortical lesions and their correlation to in vivo deficits (see Clarke and Morosan, Chapter 2) or perform invasive electrocorticographic recordings to study neural correlates of cognition (see Howard, Nourski, and Brugge, Chapter 3). Rather, detailed anatomical images, often acquired via magnetic resonance imaging (MRI), may be combined with functional information from electroencephalography (EEG) (see Alain and Winkler, Chapter 4), magnetoencephalography (MEG) (see Nagarajan, Gabriel, and Herman, Chapter 5), positron emission tomography (PET), or functional MRI (fMRI) to obtain dynamic pictures of brain function, in disordered and normal states, with reasonable spatial and temporal resolution.

T.M. Talavage (✉)
School of Electrical & Computer Engineering, Purdue University,
West Lafayette, IN 47907, USA
e-mail: tmt@purdue.edu

I.S. Johnsrude
Department of Psychology, Queen's University, Kingston, Ontario K7L 3N6, Canada
e-mail: Ingrid.johnsrude@queensu.ca

J. Gonzalez-Castillo
Section on Functional Imaging Methods, Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, Bethesda, MD 20892, USA
e-mail: javier.gonzalez-castillo@nih.gov

D. Poeppel et al. (eds.), *The Human Auditory Cortex*, Springer Handbook
of Auditory Research 43, DOI 10.1007/978-1-4614-2314-0_6,

The hemodynamic technique of fMRI has had significant impact on neuroscience because of its ability to obtain functional information in conjunction with direct localization of the signal source. Images generated while the subject is under different cognitive states may be directly compared on a location-by-location basis, with observed differences reflecting changes in the underlying physiological state of the tissue.

fMRI rests on the assumption that local changes in hemodynamics reflect changes in neural activity, as originally posited by Roy and Sherrington (1890). The brain requires a nearly constant supply of oxygen and glucose to function properly. Although the relationship is complex and not yet fully characterized, activity in neurons is closely linked to glucose consumption, oxygen consumption, and blood flow. Although the brain accounts for only about 2% of body mass, it consumes about 20% of the body's glucose and oxygen, and receives about 20% of its blood supply (Gjedde & Marrett, 2001). By mechanisms described in the text that follows, the blood flow to a region increases when local neurons become active, also increasing blood oxygenation.

fMRI is a versatile tool that is used to characterize regional changes in cognitive state, and patterns of functional correlation, throughout the brain. fMRI measures regional changes in blood oxygenation as a proxy for neural activity. The completely noninvasive nature of fMRI permits acquisition of a large number of measures without any risk for subjects, permitting longitudinal studies in individuals. One consequence is that many longer-term dynamic cognitive processes, such as learning and plasticity, can be examined effectively (Herdener et al., 2010). An additional advantage of fMRI is that localization of activity rests solely on assumptions of regional coupling of hemodynamics with neural activity, and not on source modeling approaches (see Alain and Winkler, Chapter 4, for EEG; and Nagarajan, Gabriel, and Herman, Chapter, 5 for MEG). The spatial resolution (i.e., the minimum distance between two features before they are distinguishable) of fMRI as typically implemented in cognitive neuroscience studies ranges between 1 and 15 mm, with fMRI typically conducted using spatial resolutions finer than 5 mm. The temporal resolution is generally on the order of seconds, although with fMRI it is possible to distinguish events that are separated by a few hundred milliseconds (Menon & Kim, 1999), with advances in highly parallel MRI systems permitting observation of events with even finer resolution (Lin et al., 2004).

fMRI has been used to study a vast variety of phenomena, including auditory and speech perception, advancing understanding of the functional organization of the human brain. Findings to date represent only a fraction of the potential application of such techniques, particularly given significant advances taking place in the hardware, data processing, and interpretation of results in the rapidly growing field of cognitive neuroscience. In fact, fMRI is moving beyond its current role as largely a research tool, as new clinical applications are developed (e.g., Baciu et al., 2005; Chakraborty & McEvoy, 2008).

This chapter does not attempt to provide an exhaustive review of fMRI studies of auditory function. Rather, it introduces the hemodynamic method of fMRI as used to study diverse and novel problems in the field of auditory cognitive neuroscience.

6.2 Functional Magnetic Resonance Imaging

MRI is the general name for a broad class of techniques that exploit static and dynamic electromagnetic fields to generate images of the structure and function of biological systems. During the late 1980s, researchers began exploring the potential to use MRI (Lauterbur, 1973), then a relatively new clinical tool, to obtain functional (i.e., dynamic) images of the brain. Rapid imaging techniques and hardware originally developed for imaging of the cardiac cycle were exploited to obtain sequences of images of the anatomy in the brain with a temporal resolution of seconds. Initial functional imaging of cortex was achieved using exogenous contrast agents that permitted assessment of blood flow (Belliveau et al., 1990, 1991). At the same time, however, decoupling between oxygen consumption and blood flow was also observed in MR images using an acquisition process that was sensitive to the oxygenation level of the venous blood supply (Ogawa et al., 1990; Kwong et al., 1992; Ogawa et al., 1992). Signal intensity fluctuations associated with this blood oxygenation–level dependent (BOLD) effect have subsequently become the primary tool for functional neuroimaging. Some limitations of MRI, particularly its application to auditory fMRI, are summarized in Table 6.1 and described in detail later.

Table 6.1 Limitations of functional magnetic resonance imaging (fMRI) methods in auditory research, and associated considerations for experimental design, analysis and interpretation.

Limitation	Research considerations
BOLD is an indirect measure of neural activity.	The hemodynamic response (BOLD) lags neural activity substantially, occurring over seconds. A large component of BOLD signal in venules; location of BOLD signal may not coincide with location of neurally active tissue.
MR acquisition sequences sensitive to BOLD are necessarily also sensitive to any magnetic field inhomogeneity.	Images can be distorted, and there can be reduced or absent signal (susceptibility-related dropout) in brain regions near bone, sinuses.
MR imaging involves use of magnetic fields.	Cannot scan volunteers with ferromagnetic implants such as pacemakers, pins to repair broken bones, and many types of cochlear implant. MR imaging is not compatible with the use of hearing aids. Special MR-compatible stimulus delivery and response collection equipment must be used.
MR imaging imposes physical constraints on the volunteer.	Subject is supine, must remain very still, and is in an enclosed space. Contact with the experimenter is sporadic and via intercom system and/or video link.
MR image acquisition is acoustically noisy.	Special equipment and methods must be used in auditory research to minimize direct interference (masking) from imaging acoustic noise. The auditory pathway responds both to the desired stimuli and the imaging acoustic noise, reducing sensitivity and available dynamic range.

6.2.1 Basis of Technique

Typical BOLD–fMRI data acquisition exploits an image weighting that is sensitive to changes in the local magnetic field homogeneity resulting from changes in blood oxygenation levels (review: Norris, 2006). Fluctuations in image intensity on the order of a few percent of baseline may be observed over repeated image acquisitions, where these acquisitions are generally taken at regular intervals (e.g., 200–4000 ms). The assumption inherent in fMRI is that measured BOLD signal changes are closely associated with changes in local neural activity, and depend on the cognitive state of the subject being imaged.

The observed signal fluctuations reflect changes in the oxygenation state of hemoglobin within an image volume element (voxel). Hemoglobin exhibits a small diamagnetic moment when it is oxygenated (HbO). This moment does not appreciably alter the net magnetic field in the immediate vicinity of the HbO. However, when deoxygenated (i.e., "reduced"; HbR), hemoglobin molecules exhibit a strong paramagnetic moment that reduces local magnetic field homogeneity. Therefore, as the relative proportion of HbO increases (in the ratio [HbO]/[HbR]), the field homogeneity increases, as does the observed MR signal (Ogawa et al., 1990).

6.2.1.1 The Physiological Basis of the BOLD Signal

The hemodynamic response (HDR) measured by BOLD–fMRI (see Fig. 6.1A) reflects a complex interaction among factors—including neuronal and glial physiology, rate of oxygen consumption, blood flow, and blood volume—on a spatial scale that integrates across the hundreds of thousands of cells in each fMRI voxel. A large literature is devoted to understanding the various components of the BOLD response, and how it relates to neural activity (see Aubert & Costalat, 2002; Logothetis, 2008). Here a simplified "consensus" view of the major components of the HDR is provided. It comprises three phases.

First, increased metabolic demand associated with neural activity initially drives down [HbO]/[HbR] through increased extraction of oxygen, resulting in a small signal drop that is observed as the "initial dip" (Hennig et al., 1994). This initial dip appears to be more spatially localized to the neurally active tissue than the subsequent, larger, peak of the response, but it is transient and of low amplitude, and can be difficult to measure, making it a less generally useful feature for fMRI studies (Sheth et al., 2005).

Subsequent to the initial dip, there is a rapid increase in local blood volume, likely effected at the level of the capillaries (Malonek et al., 1997) and in the venous vasculature (Malonek & Grinvald, 1996), associated with an increase in blood flow. The delivery of oxygenated blood exceeds the increased metabolic requirements of active tissue, producing an increase in [HbO]/[HbR] and the associated image signal. The peak of this response occurs 4–7 s subsequent to the onset of the driving stimulus/cognitive state, varying by position within the brain (Saad et al., 2001).

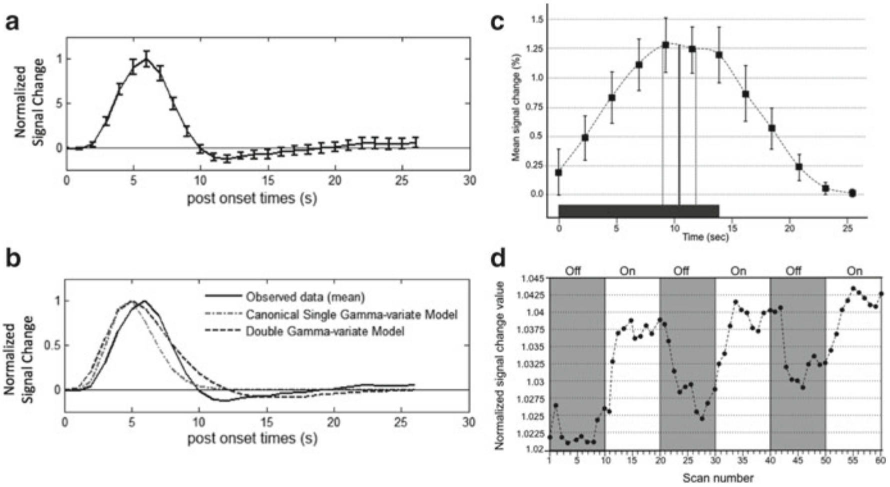


Fig. 6.1 Blood oxygenation level–dependent (BOLD) effect-based hemodynamic responses and associated models for detection and estimation. (**a**) Hemodynamic response (HDR) estimate derived from presentation of a brief (1.0 s duration) music stimulus, presented relative to stimulus onset. Note that the HDR does not break from baseline until more than 1 s after stimulus onset, requires several seconds to achieve its peak value, and equally long to resolve back to baseline, after which there is a period of undershoot relative to that baseline. (**b**) Two common hemodynamic response function (HRF) models used to detect or estimate HDRs are the gamma variate (Boynton et al., 1996) and the double gamma variate (Glover, 1999). The depicted fits were estimated using a least-squares procedure to achieve a fit to the (normalized) peak of the actual HDR depicted at top. (**c**) HDR obtained from a blocked paradigm (after Hall et al., 1999). Averaged HDR (7 subjects; TR = 2.33 s; error bars = ±1 SD) obtained from primary and secondary auditory cortex in response to a blocked paradigm involving 14 s of continuous speech followed by 14 s of no stimulus presentation. The solid vertical line indicates the peak hemodynamic response, with dotted vertical lines indicating intersubject hemodynamic lag variability (±1 SD). (**d**) Illustration of signal drift over the course of an experiment (after Grady et al., 1997). Plot is of average HDR obtained from voxels in left auditory cortex (one subject) responding to a 30/30 s "on/off" blocked paradigm of word presentations. Note that the mean activation in the last "off" block is higher than during the first "off" block, and approaches the level present in the initial "on" period

Third, when neural activity returns to baseline, the decrease in blood volume is delayed relative to the decrease in local blood flow (Mandeville et al., 1998), resulting in a drop of [HbO]/[HbR] below that observed before onset of increased neural activity. This drop leads to a prolonged "undershoot" in the observed MR signal time course, persisting for 15–25 s after the offset of the driving stimulus/cognitive state (Le et al., 2001).

The positive response just described is the most common marker of neuronal activity in BOLD fMRI, albeit a highly indirect one. Nevertheless, it is worth mentioning that task-locked BOLD signal changes with other temporal signatures can also be readily observed across the brain. These include negative-going BOLD responses (Shmuel et al., 2002, 2006) and transient responses time-locked to stimulus onset/offset (Harms & Melcher, 2003). The physiological basis of these less conventional responses is not completely understood, but recent evidence suggests that differences in neuronal processing across regions could be inferred from examination of these additional response types (Harms & Melcher, 2003; Uludag et al., 2009).

6.2.1.2 Relationship of the BOLD Signal to Neural Activity

What aspect of neuronal activity does the BOLD signal most directly reflect? First, it is likely to reflect activity over populations of cells, as the BOLD signal is measured in voxels with a typical volume of 27–64 mm^3, containing millions of neurons (Logothetis, 2008). Electrophysiologists distinguish between graded potentials, which are continuous, and spiking activity, which is all-or-nothing. At the population level, the local field potential (LFP) reflects graded (inhibitory and excitatory) postsynaptic potentials (PSPs) on a cell's dendrites (i.e., input to the cell), as well as integrative activity at the cell soma. In general, the integration across all PSPs determines whether firing by that cell will be potentiated or suppressed. Multiunit activity (MUA) reflects the spiking output of neurons. Because excitatory PSPs lead to spiking, one might expect the LFP and MUA measures to be correlated and indeed they are, but this correlation is not strong, because inhibitory PSPs may contribute positively to the LFP but negatively to the MUA. Logothetis and colleagues (Logothetis et al., 2001; Goense & Logothetis, 2008) simultaneously recorded fMRI and electrophysiological data from macaque monkey visual cortex, and observed that the LFP was a better predictor of BOLD signal change than was MUA—whether the monkeys were anesthetized or alert. This observation suggests that BOLD activity primarily reflects inputs to, and processing within, the region in which signal change is detected, rather than spiking outputs (Viswanathan & Freeman, 2007). This is a potentially vexing interpretation from the perspective of the neuroscientist, as it limits most reports of fMRI experiments to observation of increased activity, but does not allow for direct inference of increased spike activity in the associated neurons. In addition, it leads to the important implication that BOLD activity in (as a general example) area *Y* may be as reflective of neural computation in area *X* from which *Y* receives projections, as of neural computation in *Y* itself.

6.2.2 Acquisition of Data

The process of acquiring BOLD–fMRI data is achieved using imaging techniques that are sensitive to (i.e., weighted toward) magnetic field homogeneity changes taking place in and around the vasculature. The most common means to combine both reasonably high resolution and rapid temporal sampling in BOLD imaging is through the use of echoplanar imaging (EPI) (Mansfield, 1977). It permits acquisition of signal from volumes over a period of tens of milliseconds, enabling volumes as large as the whole brain to be sampled approximately every second, with relatively high resolution—typically voxels of a few millimeters per side.

6.2.3 Spatial Localization and Resolution

A critical question when using any hemodynamic imaging technique is the extent to which the observed localization of signal changes reflects the underlying locations at which neural activity has been altered by stimulus presentation or task demands. This is particularly important when experimental questions relate to precise anatomical regions: for example, how certain can the researcher be that an observed focus of activity is in the medial geniculate (and thus associated with audition) rather than in the physically adjacent lateral geniculate (and thus associated with vision)? To answer this type of question it is necessary to understand some of the factors that affect the precision with which BOLD signal can be linked to discrete populations of active neurons.

A first obvious limitation is the voxel size in the volumes acquired. Voxel volumes are typically 27–64 mm^3, and although they can be smaller, this comes at the cost of either a more powerful gradient or static fields (e.g., a 7 Tesla system, instead of 3 Tesla), or increased acquisition time and decreased detection power.

Second, although the area exhibiting a measurable HDR is approximately concordant with the area exhibiting neuronal activity (Disbrow et al., 2000; Logothetis et al., 2001), the two are not identical. The BOLD signal arises primarily in the venules—the small postcapillary component of the venous system—but the BOLD effect at moderate field strengths (e.g., 3 Tesla and below) is weighted toward larger vessels (Boxerman et al., 1995). This mismatch can result in prominent activation foci up to several millimeters from active neural tissue (e.g., Krings et al., 2001) with this shift primarily being "downstream" toward the veins. Procedures exist to reduce signal from intravascular space when imaging at low field strengths (e.g., Glover et al., 1996), although this can result in reduced contrast between metabolic states. At higher field strengths (e.g., 4 Tesla or greater) the localization of activation improves, because the contrast is better localized to the extravascular space (Gati et al., 1997).

Finally, data processing affects the effective spatial resolution of fMRI, because the pre-processing steps required for single-subject and group analysis (e.g., resampling after spatial normalization and/or spatial smoothing) result in averaging of signal across voxels, which degrades spatial resolution. Data processing is discussed in detail in Section 6.4.

In summary, the precision with which activation sites can be localized or segregated depends on factors that span both the underlying physiology and decisions about how to acquire and process the data. For general reference, typical resolution in group studies in the cognitive neuroscience literature can range between 5 and 15 mm, but acquisitions can be optimized to yield resolution of 2 mm or less (e.g., Schonwiesner & Zatorre, 2009).

6.2.4 Temporal Resolution

The temporal resolution of fMRI is limited by two factors: the rate at which measurements are obtained and the sluggishness of the hemodynamic response (HDR). Samples of a single slice may be acquired as rapidly as every 20–40 ms, or whole brain data may be acquired every second (or more slowly). A typical acquisition would comprise 30–40 slices, each 2–5 mm thick, over a 1- to 3-s time window, with each slice acquisition taking place in 10–100 ms, dependent on field strength, number of coils, and acceleration factor (Pruessmann et al., 1999; Lin et al., 2004). The sampling rate, i.e., time between successive acquisitions of the same slice, also known as the repetition time (TR), serves as the primary limitation, but although the slow rate of change observed for the HDR is such that sampling every 1–2 s is typically sufficient to accurately characterize the response amplitude to a given stimulus, it remains possible to detect variations in response times of less than 100 ms (Saad et al., 2001).

6.2.5 Signal Artifacts

fMRI techniques are intentionally sensitive to subtle fluctuations in local magnetic field strength, but magnetic field inhomogeneities that are not related to neuronal activity represent a substantial source of noise in fMRI experiments. Whereas protons in tissue (such as brain) are relatively fixed and can produce a substantial local magnetic moment, the rapid, unconstrained movements of protons in air-filled cavities result in a minimal local magnetic field. Boundary conditions dictate that the magnetic field must be continuous across interfaces between these two types of spaces, creating steep gradients between tissues and voids that can adversely affect image quality, sometimes resulting in complete loss of signal from affected regions. Areas particularly prone to these “susceptibility artifacts” include inferior frontal and inferior temporal cortex (Ojemann et al., 1997). “Dynamic” variations that can produce signal artifacts arise from respiration and cardiac fluctuations, both of which can produce changes in the distribution of the mass of the body without changing that mass. These variations in density and distribution lead to time-varying shifts in boundary conditions between the body and the static magnetic field of the MRI, with associated signal fluctuations throughout (Frank et al., 2001). Signal variations brought about by these fluctuations are considered in Section 6.4.1.

6.2.6 Considerations for Auditory fMRI Experiments

From the perspective of a researcher interested in the auditory system, the MRI environment is quite problematic (e.g., see Table 6.1). The following sections address the particular acoustic confounds associated with fMRI and current “best practices” to obviate these effects.

6.2.6.1 Imaging-Related Acoustic Noise

The majority of EPI techniques for fMRI involve switching of gradient fields at audible frequencies (20 Hz–20 kHz), yielding an intense sound at the time of acquisition of each slice in the volume. The acoustic noise associated with typical EPI sequences is of high frequency, with the majority of energy between 0.5 and 2 kHz, and significant harmonics to above 10 kHz. The sound is especially intense in the bore of the magnet, at the subject's head, generally peaking between 94 and 135 dB SPL (Foster et al., 2000; Ravicz et al., 2000).

Acoustic noise associated with image acquisition is problematic in fMRI experiments for several reasons. First, it can be distracting, changing patterns of behavior and patterns of activation when it is present (e.g., Tomasi et al., 2005). Second, it can mask experimental auditory stimuli, making them harder to hear, and imposing a requirement for auditory perceptual segregation of target and masker. Clearly, this is undesirable for auditory studies. The perceptual challenges imposed by this acoustic noise can interact with the experimental task and with the subject group under study (e.g., a group with an auditory perceptual impairment) in unexpected and hard-to-control ways.

The generated acoustic noise is also of concern because it represents a generally undesired stimulus that can produce appreciable responses throughout auditory (Bandettini et al., 1998) and nonauditory (Zhang et al., 2005) cortices. The response of auditory cortex to the acoustic noise can result in a decrease in observed percent signal change, associated statistical values, and the number of significantly active voxels in auditory cortex arising from experimental stimuli (e.g., Olulade et al., 2011). The described effects of the impulse-like acoustic noise are likely to be greatest in primary auditory cortex, which exhibits particular sensitivity to sound onsets (Giraud et al., 2000; Harms & Melcher, 2002). Further, the HDR evoked by the experimental manipulation adds nonlinearly to the HDR induced by the acoustic noise, presumably due to variable levels of saturation of the hemodynamic system (2004). As a result, simply "subtracting" the HDR associated with the acoustic noise is not possible.

6.2.6.2 Reduction of Imaging-Related Acoustic Noise

Mitigation of confounds related to imaging-related acoustic noise is generally achieved by means of passive attenuation. Subjects are typically given ear plugs, and/or circumaural ear protectors. Such a combination can yield 25–40 dB of attenuation, depending on frequency and type of protection used (Ravicz & Melcher, 2001; Hall et al., 2009). Such measures must be compatible with the MRI environment (e.g., no ferromagnetic materials) and are also constrained by the physical space limitations associated with the equipment used for data acquisition. For example, the hardware required to image the head may not provide sufficient space for circumaural headphones that maximize attenuation of the noise. Recent efforts have demonstrated that active noise cancellation is possible, with reports of single-frequency attenuation of up to 35 dB (Mechefske et al., 2001; Hall et al., 2009).

The method chosen to attenuate acoustic noise must be compatible with sound stimulus delivery. Several MR-compatible sound delivery systems are available, including pneumatic systems where the signal is propagated through air-filled tubes to the subject in the scanner, and presented through insert earplugs; and wired systems, in which the delivery electronics are optically isolated within the scanner room. Some electrostatic headphone systems have been developed that are MR-compatible, permitting high-fidelity presentation of auditory stimuli. Earplug versions of these electrostatic systems can sometimes (if the hardware for head imaging offers enough space) be used with circumaural ear defenders for further attenuation. The frequency response characteristics of any system should be carefully evaluated, as should the sound attenuation afforded, and the potential for the equipment to produce image artifacts.

6.2.6.3 Image Acquisition for Auditory Experiments

In the absence of full attenuation, imaging-related acoustic noise continues to make fMRI studies of the auditory system challenging. Several methods have been developed to compensate for this problem. Most of these methods introduce compromises to the speed at which data are acquired, to brain coverage, or to both (Loenneker et al., 2001). The most common solution (see Fig. 6.2B) takes advantage of the 4- to 7-s delay between HDR onset and peak to intersperse volume acquisitions with periods (of up to a few seconds duration) in which no images are acquired. Because the primary confounding noise is generated by the MR system only during acquisition, this provides a quiet interval in which to present stimuli (Eden et al., 1999). The hallmark of these clustered volume acquisition (CVA) or "sparse" sampling techniques is that the TR (i.e., the time from volume acquisition onset to volume acquisition onset, including the gap) of the imaging sequence is notably longer than the time expended to acquire a complete brain volume (Edmister et al., 1999; Hall et al., 1999). Sequence timing is generally chosen such that the peak of the stimulus-induced HDR coincides with data acquisition. Another advantage of CVA/sparse sampling is that a long TR (e.g., 3 s or greater) allows the longitudinal magnetization to recover almost completely, so that the dynamic range of image signal is nearly maximal for every acquired volume, increasing the signal-to-noise ratio (SNR) relative to volumes acquired with a shorter, more conventional TR (Edmister et al., 1999; Hall et al., 1999).

The CVA/sparse sampling technique is powerful, but has limitations because of the long gap between acquisitions. This gap is generally chosen to be longer than the longest auditory stimulus to be presented in an experiment. In the case of short

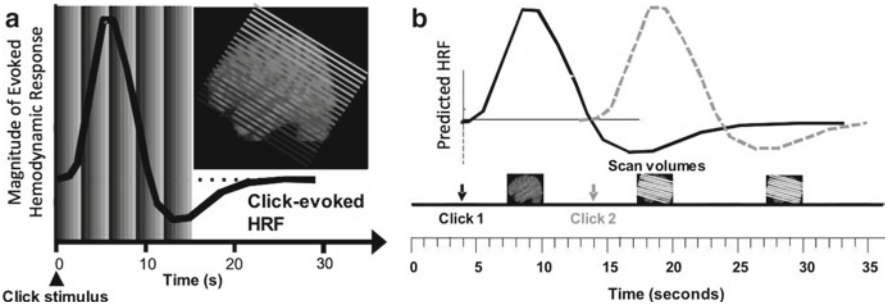


Fig. 6.2 (**a**) Illustration of acquisition of a typical brain volume and its timing, relative to the hemodynamic response (HDR). fMRI data are typically acquired one brain slice at a time, with a whole volume comprising multiple slices (here, 21) obtained over 1–3 s (here, 3 s). As a result, each slice is measured at a different point relative to a transient impulsive response (e.g., to a click). Note that for the acquisition displayed here, the middle of the brain (around auditory cortex) is not scanned while the evoked HDR is at its peak. (**b**) An illustration of an experimental design using clustered volume acquisition (CVA)/sparse imaging techniques for an event-related paradigm (Eden et al., 1999; Edmister et al., 1999; Hall et al., 1999). The stimulus is presented during periods in which no imaging data are acquired (i.e., when no imaging-related acoustic noise is present), with the resulting HDR expected to achieve its peak value near the time of volume acquisition. Such techniques are intended to maximize the peak HDR amplitude, and are well suited to both detection and estimation

stimuli (i.e., single words) the TR may be kept relatively brief (e.g., 2.5 s in the case of presentation of short words; see Orfanidou et al., 2006), allowing the HDR to be observed at multiple post-onset times and preserving statistical power when using regression-based analyses, relative to continuous imaging (i.e., no gap is inserted between volume acquisitions, so slices are acquired throughout the TR period; see e.g., Edmister et al., 1999). However, when this gap becomes relatively long (e.g., 5 s or greater), not only are fewer measurements made during the experiment, but typically the HDR is sampled at only one time (ideally on or very near the peak; but see Belin et al., 1999). This becomes most problematic whenever estimation of the shape and timing (e.g., time of onset, peak or return-to-baseline) of the HDR is desired (see Section 6.3.3.3). Several techniques have been developed that permit sampling of the response with high temporal resolution, while still permitting delivery of stimuli in the absence of the imaging-related acoustic noise (Schmithorst & Holland, 2004a; Schwarzbauer et al., 2006).

As imaging hardware has advanced, it has become feasible to develop image acquisition sequences that appreciably reduce acoustic noise while preserving an acceptable SNR. Techniques that may prove widely beneficial for auditory studies include use of continuous acquisition protocols that have a reduced impact on primary auditory cortex (Seifritz et al., 2006), or acquisition protocols where the parameters are specially chosen to result in reductions in the imaging acoustic noise of 20 dB or greater (Peelle et al., 2010).

6.3 Design of Hemodynamic Imaging Experiments

Hemodynamic imaging methods can be used for three broad classes of research inquiries: (1) they can reveal which particular brain areas are sensitive to stimulus and/or task demands (*detection* of hemodynamic response change); (2) within such areas, they can reveal subtle changes in the shape or timing of the hemodynamic response (*estimation* of hemodynamic response change); and (3) they can reveal functional coupling among discrete regions related to stimulus or task demands.

6.3.1 Experimental Constraints

Before developing an experiment, it is critical to consider the environment of the imaging suite when developing imaging studies. In addition to the acoustic noise of the MR system, which can distract and annoy participants and mask experimental stimuli, the experimental setup is of particular import: the participant is generally tested while supine on a bed, encircled by an MR scanner bore, which is typically narrow (55–65 cm in diameter) and long (1.0–1.9 m), increasing potential for claustrophobia, and restricting movement. In addition, participants are typically imaged while their head is enclosed in hardware that limits their ability to perform actions and interact with the researcher. All equipment used in MR studies, such as stimulus presentation and response collection equipment, must be compatible with intense magnetic fields (note that 1 Tesla is approximately 20,000 times as strong as Earth's magnetic field). MR acquisition is highly sensitive to motion and participants must remain as still as possible. Behavioral responses during imaging studies are often limited to button presses for this reason. Finally, issues related to patient comfort—e.g., effort exerted in holding still, fatigue related to continual presence of imaging-related acoustic noise—limit typical fMRI sessions to about 1 hour, usually divided into multiple runs of several minutes apiece, allowing for regular contact with the participant during breaks between runs.

6.3.2 Experimental Design Principles

Because hemodynamic imaging experiments rely upon a contrast between two or more conditions, they are often conceived as simple subtraction designs, in which activity levels evoked by conditions are compared. Typically one condition serves as a low-level "baseline" condition, and conditions are matched as closely as possible on all features except for the independent variable. A popular way to achieve this is to hold the stimulus constant while changing the task, or to change the stimulus while holding the task constant. The latter approach has been, and continues to be, deservedly popular for auditory and speech perceptual studies (e.g., Binder et al., 1994).

Subtractive designs often rely on an assumption of “pure insertion.” This is the assumption that the higher-level—experimental—condition (e.g., hearing speech) incorporates all cognitive and perceptual aspects of the lower-level—control—condition (e.g., hearing noise, matched in spectrum and envelope to speech), with an extra “inserted” feature of interest. Such assumptions may not always be justified in cognitive experiments because conditions can differ in ways that are not obvious to the experimenter, and cognitive components may not “add” linearly to each other (Friston et al., 1996). For example, an individual may be engaged in many active cognitive processes during so-called “rest” or other “baseline” conditions (see Binder et al., 1999), which may result in false negatives when this condition is used as the basis for comparison.

There are two popular ways to avoid violations of the assumption of pure insertion. One approach is to use a parametric design. Unlike subtraction designs, in which conditions differ qualitatively (e.g., a feature of interest is either present or absent), parametric designs treat the feature of interest quantitatively, as a continuous variable, and many different levels (amounts or intensities) are tested. The assumption underlying parametric designs is that the magnitude of the HDR reflects sensory or cognitive “load,” such that the greater the processing engaged by a task, the greater the magnitude of the BOLD signal change. Such designs provide something like a “dose–response curve” for the stimulus or task parameter. For example, several researchers have investigated how variations in rate of word presentation are reflected in activity in human auditory cortical regions (e.g., Binder et al., 1994; Buchel et al., 1998). Parametric studies are also useful for determining whether HDRs arising from stimulation are linear, nonlinear, whether they vary in different regions of the brain, and whether they are specifically relevant to a particular sensory or cognitive function (Büchel et al., 1998).

Another approach to the subtraction problem is to look for generalization of the experimental effect across multiple subtractions, as a potential confound might apply to one subtraction, but is unlikely to apply to multiple, independent, subtractions. This can be accomplished using a factorial design, in which two or more independent variables are “crossed.” In a “crossed” design, the effect of each level of one independent variable (e.g., type of acoustic degradation applied to speech) is tested at multiple levels of another independent variable (e.g., multiple levels of intelligibility). Factorial designs also allow for the testing of interactions, which permit more specific inferences regarding the perceptual or cognitive locus of an observed effect (Henson, 2006a). Within the factorial framework, *main effects* are the average effect of one factor, collapsed across all the levels of other factors (e.g., main effect of intelligibility, collapsed across type of acoustic degradation). The examination of such main effects can be supplemented by *conjunction analyses* (e.g., see Henson, 2006b) that assess replicability across multiple, independent contrasts, for example, the extent to which the relationship between BOLD signal and intelligibility is similar for different types of acoustic degradation.

To provide a more complete illustration of parametric and factorial approaches, consider the study of sentence intelligibility by Davis and Johnsrude (2003), which employed a parametric manipulation of intelligibility within a factorial design.

In this study, spoken sentences were degraded using three acoustically distinct manipulations, and each degradation type was applied to three different degrees, yielding nine different levels of intelligibility. Extensive regions of bilateral temporal cortex exhibited a significant linear correlation with increasing speech intelligibility (collapsed across types of degradation—i.e., a main effect), suggesting that these areas are sensitive to the amount of speech understood. In contrast, other frontal and temporal-lobe regions exhibited an inverted U-shaped relationship with intelligibility, suggesting that such regions are sensitive to the amount of effort required to understand the degraded speech. Davis and Johnsrude (2003) observed that a subset of the regions exhibiting a positive linear correlation with intelligibility, around auditory cortex bilaterally, also showed a differential response to the three forms of distortion (i.e., an interaction between the two factors) consistent with sound-form–based processing. A conjunction analysis revealed that a different subset of areas, within the superior and middle temporal gyri, hippocampus, and left inferior frontal lobe, were insensitive to the acoustic form of sentences, suggesting more abstract, nonacoustic, processes.

A limitation of factorial designs is that, with every added factor, the number of conditions increases multiplicatively. A 2 × 2 factorial design is 4 conditions, and a 3 × 3 is 9. But add another 2-level factor to that 3 × 3 design, and suddenly you have 18 conditions! How many conditions are practical? In fMRI studies, the maximum number of conditions that can be tested in any single experiment is about 12–16. The number is limited for several reasons. First, the amount of time for which a person can be scanned is limited by practical constraints (see Section 6.3.1) and conditions need to be replicated to increase the contrast-to-noise ratio sufficiently to observe the experimental effect. The number of replications depends on a variety of factors including the size of the experimental effect, and the strength of the static magnetic field in the MR system, but, generally, the more replications the better (see Henson, 2006b). Also, conditions that will be contrasted need to be tested quite close together in time. fMRI data include substantial low-frequency fluctuations arising from multiple sources, including distorted representations of physiological noise (e.g., heartbeat and respiration; Lowe et al., 1998). See Figure 6.1D for an example of low-frequency "drift" over time (Grady et al., 1997). Although drift correction and other temporal processing mechanisms can reduce these effects (see Section 6.4.1), there are limits to the efficacy of such corrections. If stimuli or events to be compared are too far apart in time, then the evoked signal changes for these events will be at low frequencies, and may be indistinguishable from noise. For similar reasons it is best to have relatively few conditions, because the more conditions that are included, the farther apart in time (on average) will be any two conditions to be contrasted. Further, given practical time constraints on experiment and session durations, increasing the number of conditions to be tested necessarily leads to fewer replications of each condition, and thus to fewer measurements with which to generate statistical power.

6.3.3 Experimental Designs

To best choose a design, the researcher must clarify two issues. The first is the potential sensitivity of the perceptual or cognitive processes of interest to scanner noise. The second is the experimental objective: whether it is primarily (1) detection of activity related to a relatively static or sustained (perceptual or cognitive) state; (2) detection of activity related to a more transient perceptual or cognitive process; or (3) estimation of the time course of the response. (Of course, there are also some questions that may be best answered using a method other than fMRI).

The potential sensitivity of the processes of interest to scanner noise will determine whether a researcher chooses to use CVA/sparse imaging (see Section 6.2.6.3) or continuous imaging. Continuous-imaging approaches (see Edmister et al., 1999) are typical in other domains of cognitive neuroscience research, and certainly yield the most data and the best temporal resolution, but they are not recommended if the experimental question concerns regions like auditory cortex that will be saturated by imaging-related acoustic noise, or where the experimental stimulus or task will be fundamentally changed by the continuous presence of such high-intensity background noise.

If the objective is detection, many different designs are possible. When the perceptual or cognitive processes of interest are relatively sustained over time, then it is appropriate to use a "blocked design," in which stimulus presentation is blocked by condition. Such designs are discussed in Section 6.3.3.1. If the perceptual or cognitive process is more transient in nature, then an "event-related design" in which stimuli from different conditions are intermingled and presented in rapid succession is appropriate. Such designs are considered in Section 6.3.3.2. Both of these designs may be used with continuous acquisitions or CVA/sparse sampling techniques. Estimation experiments, which by definition are aimed at modeling transient events with good temporal precision, would generally use continuous imaging and event-related designs, as discussed in Section 6.3.3.3. No general design considerations pertain to experiments investigating connectivity; the relevant analyses, and other, multivariate and data-driven, analyses, are discussed in Section 6.4.6.

6.3.3.1 Designs for Detection: Measuring State-Related Perceptual and Cognitive Activity

Often, the experimental question concerns where in the brain (either focally, or in a distributed fashion) neural tissue sensitive to perceptual and/or cognitive activity is found. Further, it may be reasonable to consider the perceptual and cognitive processes of interest as "states," extended in time, and not as fundamentally transient processes. Indeed, many sensory, perceptual, and cognitive processes can be considered to evoke a "steady-state" brain response (e.g., listening to amplitude-modulated sounds). For such studies a "blocked" design is appropriate. In blocked designs, each block of a single condition (e.g., stimulus presentation, task, or cognitive state) is initiated and sustained for 10–30 s, while the brain is imaged, typically continuously (i.e., without the gap characteristic of CVA/sparse designs). Henson (2006b) suggests that an optimal length for a single condition in a blocked design is 16 s, as long as this is compatible with psychological considerations (e.g., Will participants get confused if tasks switch too quickly? Is 16 s long enough to complete a "trial"?). Birn et al. (2002) documented that for blocked designs contrasting two experimental conditions, detection is best achieved using a 50% duty cycle—that is, if 16 s is the best duration for the presentation of a state, then 16 s is also the optimal duration for that state to be absent. If imaging-related acoustic noise is an issue, then a block design can be approximated using CVA/sparse sampling (Section 6.2.6.3), with a relatively long silent period in which a stimulus or sequence of stimuli is presented; typical TRs in such studies would be 9–10 s or more (i.e., a 7 s gap for steady-state stimulus perception, followed by a volume acquisition).

Blocked-design experiments are straightforward to analyze and do not require detailed consideration of the temporal characteristics of the HDR. A block of stimuli is considered to be like a long stimulus, with a duration equal to the block length, resulting in an amplified HDR with a plateau that extends in time (see Fig. 6.1C). A CVA/sparse sampling procedure would be designed to catch the HDR at the plateau, whereas continuous sampling would capture the entire response, most of which (for reasonably long blocks) is the plateau phase.

The disadvantage of blocked designs is that they integrate activity over many seconds. The activity measured during these acquisitions reflects aggregate perceptual, cognitive, and motor processes that are occurring over that time period, making interpretation of activity difficult. Some psychological phenomena cannot be considered to be sustained "states." For example, to study the activation patterns related to different percepts of an ambiguous stimulus, the responses to each presentation must be sorted according to which percept the subject was experiencing, perhaps as indicated by a key press at a perceptual switch. Similarly, if activation related to "correct" compared to "incorrect" trials is of interest, then a block design is not appropriate. Further, the sustained/repetitive nature of blocked designs both precludes the study of cognitive processes that require novelty and can introduce anticipatory effects and/or habituation effects. In spite of these limitations, such designs have made important contributions to our knowledge of auditory and speech perception (e.g., Binder et al., 1994; Talavage et al., 2000).

6.3.3.2 Designs for Detection: Measuring Transient Perceptual and Cognitive Activity

The detection of activity related to transient perceptual and cognitive processes is best approached using an "event-related" design, in which the unit of analysis is individual "events" of relatively short duration, rather than a sustained period of event presentation (Buckner, 1998). Although most event-related designs use continuous imaging to maximize temporal resolution, if acoustic noise is an issue, then CVA/sparse sampling is also possible.

The central assumption of event-related designs is that activity will be evoked transiently and discretely by each occurrence of an experimental stimulus or task (an “event”), such that when events of different types are presented, separated in time, evoked responses specific to each type can be discerned. This is conceptually like electrophysiological experiments in which it is common to present a single instance of a stimulus or task, measure the evoked physiological response, and average over many events of the same type to observe the characteristic evoked response. In fMRI, the range of presentation rates within which it is appropriate to test for experimental effects is bounded below by the presence of low-frequency measurement and physiological noise, and above by the blurring associated with the slow temporal characteristics of the HDR (Fig. 6.1A). This range generally corresponds to interstimulus intervals of between 2 and 20 s (Bandettini & Cox, 2000), with optimal selection for accurate estimation of response amplitude and shape being a function of stimulus duration (see Birn et al., 2002). If trials of two types to be contrasted are close together in time (such that the HDR evoked by the first is not completely recovered at the time of the second) then it is important to jitter or randomize the time interval between them so that activity evoked to the first stimulus type can be distinguished from that evoked by the second (Burock et al., 1998). Such randomized “rapid-presentation” event-related experiments have been documented to work effectively in the visual system (Boynton et al., 1996), but are problematic in the auditory system because they require imaging using a short TR, with the attendant confounding effects of acoustic noise from the MR systems.

Where experimental considerations preclude the use of continuous imaging, an event-related design can be approximated using CVA/sparse sampling (Orfanidou et al., 2006; Schwarzbauer et al., 2006). One option is to use a short TR—e.g., Orfanidou et al. (2006) used a 1.1 s image acquisition interspersed with a 1.4 s gap in which single words were presented. Alternatively, special techniques in which stimulus presentation is followed by multiple, short, image acquisitions can permit researchers to distinguish among transient events occurring at different post-presentation times, as the stimulus is being perceived. Schwarzbauer et al. (2006) presented spoken sentences in 8 s gaps, each followed by a sequence of five, 1 s duration acquisitions. They were able to distinguish neural activity related to key presses occurring at the beginning of sentences, from that related to key presses occurring near the end.

6.3.3.3 Designs for Estimation of the Hemodynamic Response

Another broad class of experimental questions, for which the event-related designs presented in the previous section are particularly well suited, require *estimation* of the magnitude and timing of the BOLD response. Obtaining the actual timing of the HDR for each responsive voxel, instead of simply assuming a common hemodynamic model (see Section 6.4.4.1) permits study of how hemodynamics vary across cortical regions, individuals, and/or subject populations (e.g., healthy vs. patient), and identify subtle differences in processing roles between regions known to be recruited by particular experimental tasks.

The HDR can be highly stable in shape, amplitude, and timing within a particular brain area (Aguirre et al., 1998), provided the state of the vascular system at the time of stimulus presentation is consistent (Olulade et al., 2011). However, activation time courses are substantially more variable when compared across regions, within an individual, and across participants, within a given region (see Handwerker et al., 2004). Based on these results, it has been concluded that, with appropriate constraints (i.e., comparison within a single region in each subject), event-related fMRI can be used to resolve fine timing differences, even of less than 100 ms (e.g., Menon & Kim, 1999).

Estimation of the HDR is also desirable in adaptation (or priming) studies. In such studies, the researcher looks for an altered HDR to a target stimulus following a prime that shares one or more key features with the target, relative to a target following a prime that lacks common features. In general, when a stimulus feature is repeated, it elicits less activity from neural populations involved in representing that feature (Weigelt et al., 2008). This adaptation effect is a useful tool for probing sensitivity of brain regions to stimulus features. For example, Sammler et al. (2010) used an adaptation method to study whether lyrics and melodies are processed together or separately when hearing unfamiliar songs. They induced selective adaptation for either lyrics or melodies by manipulating the degree of repetition of the relevant component. Greater activity in a region when a component is varied, compared to when it is repeated, indicates sensitivity to that component. They found that the superior temporal gyrus was bilaterally sensitive to both components; whereas there was some indication that left anterior superior temporal sulcus was sensitive to lyrics but not melodies. Such differentiation in the response allowed them to suggest that the two components of songs are, to at least some degree, processed independently (Sammler et al., 2010).

6.4 Data Analysis

Analysis of hemodynamic imaging experiment data is generally conducted using either model-based or data-driven "exploratory" procedures, with the former preferred for the purpose of detection or estimation of the hemodynamic response in a given region of the brain, and the latter used when seeking to characterize interactions among regions or distributed patterns of activity. Imaging analyses can be complex (see Fig. 6.3) and there is no single "correct" procedure. Therefore, researchers must be vigilant to avoid sources of bias in their analyses that could invalidate their conclusions (Bennett et al., 2009; Kriegeskorte et al., 2009), with this goal most likely achieved by procedures that conform to evolving statistical "best practice."

Analysis in fMRI is conducted on an individual or group basis, typically using parametric, model based methods such as *t*-tests, regressions, or analysis of variance (ANOVA). In most experiments, researchers seek to characterize typical patterns of activity over a group of subjects, such that results can be generalized to a population.

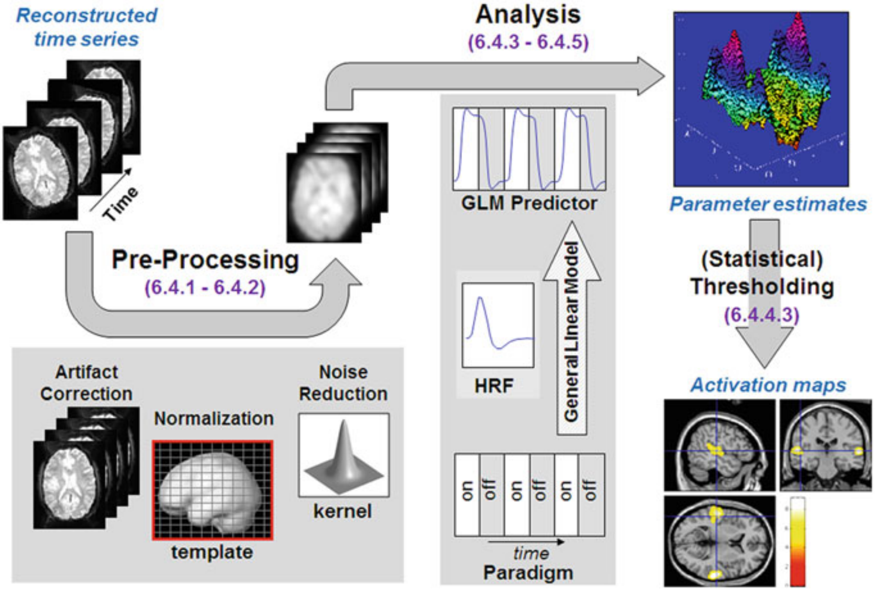


Fig. 6.3 An example data processing stream for fMRI, illustrated for the case of a blocked paradigm (Section 6.3.3.1) analyzed using a general linear model (GLM) approach (Section 6.4.4) to reveal bilateral activation in auditory cortex

Given that intersubject variability is generally greater than intrasubject variability, it is necessary to treat the subjects as a random factor in a group-level "random-effects analysis." Although such analysis makes several assumptions about the distribution of the individual subject performance relative to the group mean, it is a reasonable approach that allows generalization of results beyond the experimental sample to a larger population, and also permits comparison of experimental results with those obtained in other experiments.

Before data can be analyzed, however, the many known sources of noise must be addressed. Below, common processing steps applied to remove noise from fMRI data (see Fig. 6.3, left) are first described, followed by statistical analysis methods (see Fig. 6.3, middle and right).

6.4.1 Data Pre-Processing for Removal of Noise and Artifact

fMRI data commonly undergo several pre-processing steps to compensate for noise and artifact sources related both to the duration of image acquisition and the interaction of the subject with the magnetic environment. Typical pre-processing stages that correct for these confounds may include (1) compensation for physiological artifacts, (2) slice-timing correction, (3) compensation for subject movement, and

(4) efforts at reduction of low-frequency noise. Each of these corrections is intended to improve the ability to detect, characterize, and compare HDRs across the entire brain, within a single subject or across multiple subjects.

The presence of a human body in the bore of the magnet perturbs both the static and dynamic magnetic fields involved in acquisition of fMRI data. In addition to boundary conditions that produce susceptibility artifacts (see Section 6.2.5), temporal variations associated with respiration and the cardiac cycle produce time-varying fluctuations in the magnetic field (respiration) or physical position and deformation of structures (cardiac cycle) at remote locations within the body, including the brain (Wowk et al., 1997). Concurrent monitoring of respiration and cardiac rhythms during fMRI permits some of these fluctuations to be removed from time-series data (e.g., RETROICOR; Glover et al., 2000), providing for an increase in the SNR. Note that such compensation is critical to many connectivity studies, because cardiac and respiration signals can produce widespread correlations (Lowe et al., 1998).

When the precise temporal characteristics of the BOLD response are of interest (as in an estimation-based design; see Section 6.3.3.3), the timing of slices within an image acquisition is important to consider. Depending on the TR, several seconds might elapse between the first and last slices acquired. For a transient event (e.g., a click), each slice will necessarily be measured at a different point in the evoked HDR (see Fig. 6.2A). During experiments involving rapid volume acquisitions (e.g., TR < 1 s) or blocked paradigms, these offsets may not appreciably affect analysis. However, in model-based analysis of transient evoked responses (i.e., particularly for event-related paradigms), a fixed hemodynamic response function (HRF) model that is offset by as little as 1.5 s can substantially alter obtained statistics (Henson et al., 1999). Typical compensation approaches interpolate and subsequently resample the data to obtain a common measurement time across all slices in the volume (Calhoun et al., 2001).

Motion correction is generally essential, because, in spite of the best efforts of experimenters, no procedure—be it use of a bite block, padding, vacuum pillows, or a rigid frame—fully immobilizes alert subjects during the course of an imaging experiment without producing excessive discomfort. Given the high spatial resolution of fMRI data (typically 2–4 mm in-plane), movements of as small as 1 mm in any direction can represent a substantial change in the part of the brain associated with a given voxel. This is problematic when one of the experimental goals is to localize a source precisely. Rigid-body motion correction (e.g., all volumes registered to a common reference volume) is the most common algorithm to compensate for movement (Cox & Jesmanowicz, 1999). Other procedures may be more appropriate when only a subvolume of the brain has been acquired (Greve & Fischl, 2009).

As mentioned in Section 6.3.2, fMRI data contain appreciable low-frequency fluctuations that can confound the analysis of data for the (relatively) low-frequency fluctuations associated with the HDR (e.g., Fig. 6.1D). Temporal noise is usually reduced through drift correction and high-pass filtering (e.g., with a cutoff frequency chosen such that the experimental design does not introduce meaningful variation at lower frequencies).

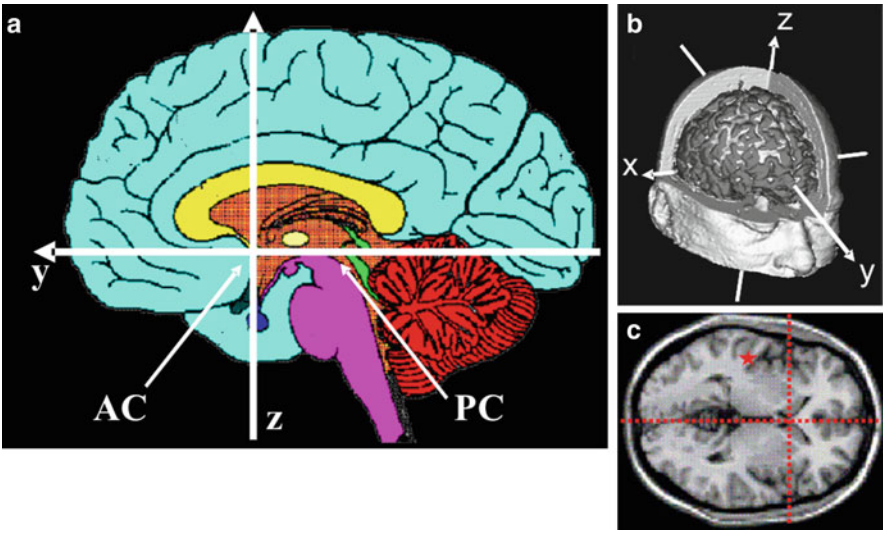


Fig. 6.4 Example of a standardized stereotactic coordinate system to which data may be spatially normalized for intersubject comparison or comparison with extant literature. (**a**) Axis origins are all at the anterior commissure (AC). (**b**) Any location in the brain can be indexed by coordinates along these three axes, typically using units of mm. (**c**) After spatial normalization, a particular set of coordinates approximately refers to the same brain region in all subjects. For example, the star is at (x, y, z) = (–50, –15, 0)—approximately 4.5 cm left of midline, 1.5 cm anterior to the anterior commissure, and 1 cm above a line connecting the AC with the posterior commissure (i.e., the AC–PC line)—describing a position in the left inferior frontal gyrus

Spatial noise is generally modeled as a spatially and temporally stationary Gaussian process. This spatial noise lends itself well to increasing SNR through spatial blurring with a (Gaussian) kernel—commonly 6–10 mm full-width-at-half-maximum—under the assumption that HDRs that extend over adjacent voxels will be reinforced, while the (white) noise will tend to zero. This spatial blurring operation also proves useful in group analyses, compensating for residual intersubject variability in anatomy.

6.4.2 Transforming Data into a Standard Reference Space

A further common pre-processing step is spatial normalization (also referred to as registration or warping) to a standardized stereotactic space. Such a space provides a means to compare activation across different brains (see Fig. 6.4) and to other studies in the literature. Any location in the brain can be indexed by values of three orthogonal axes, generally measured on a millimeter basis (Fig. 6.4C). As a general rule, the x-axis defines medial–lateral (left–right), the y-axis defines anterior–posterior (front–back), and the z-axis defines superior–inferior (up–down). In humans, the origin is usually in the vicinity of the anterior commissure.

Stereotactic coordinates are important in neuroimaging because individuals vary substantially in the location and extent of cortical folding, although major sulcal and gyral landmarks are broadly similar (Rademacher et al., 1993; Amunts et al., 1999). Therefore, conventional group analysis requires that the brain data be normalized to some common reference space to permit fixed and random effects group analyses, and for automatic identification of landmarks to which particular functions may reliably be ascribed (e.g., Morosan et al., 2001).

Regardless of the target space, spatial normalization is typically effected in an automated fashion. Common reference spaces include those associated with a brain atlas—for example, the Talairach and Tournoux atlas (Talairach, 1988), or one of the multisubject (152, 305, or 452 subjects) templates developed by the Montreal Neurological Institute (Brett et al., 2002). Some normalization routines do not require a reference template but define the atlas from the group of subjects under study, either in 3D space (Ashburner, 2007), or by surface-based averaging techniques (Fischl et al., 1999).

Normalization compensates in part for sulcal and gyral variability across subjects, with the expected consequence that overlap of functional activation among subjects will be increased. Although function is determined more by cyto- and chemoarchitecture and connectivity in the brain than by the configuration of sulci and gyri, these microanatomical characteristics, to a first approximation, align with gross anatomical landmarks such that registration of brains to a standard space using these gross landmarks results in better registration of microanatomical regions as well (Fischl et al., 2008). See (Brett et al. 2002) and (Poldrack and Wagner 2004) for reviews of normalization, atlases, and the assumptions inherent in their use.

6.4.3 General Analysis Procedures

Once the data have been pre-processed to minimize spatial and temporal noise effects, they undergo quantitative analysis. As described in Section 6.3, experimental questions can pertain to detection of responses induced by a task or stimulus, estimation of the shape and magnitude of responses induced by a task or stimulus, or to connectivity assessment of functional relationships (e.g., as evidenced by within-condition correlation) between regions of the brain and how these change depending on the task or the type of stimulus. Model-based analyses (Section 6.4.4) are still the most common type of analysis and are appropriate for many types of detection and estimation experiments. More data-driven and multivariate analyses may also be appropriate for detection experiments, particularly if the question concerns how tasks or stimulus types can be characterized in terms of distributed patterns of activity across a region (or the whole brain). These are discussed in Section 6.4.5. Connectivity analyses (Section 6.4.6) can be conducted in many different ways, from model-driven to data-driven, univariate or multivariate analyses.

6.4.4 Model-Based Analysis

The most popular type of analysis uses a general-linear-model (GLM) approach to evaluate regional sensitivity to stimulus or task demands (Friston et al., 1995; Worsley & Friston, 1995). First, predictor variables are created based on the times at which experimental conditions were presented throughout the experiment (see Fig. 6.3). Then, the degree to which these predictor variables fit the data at each voxel (or in each region of interest) is determined by correlation. When more than one predictor variable (i.e., experimental effect) is present, the observed time series in each brain region or voxel is considered to reflect an optimally weighted average of the predictors, such that residual error is minimized.

The GLM and other model-based procedures are predicated on the assumption that the hemodynamic system—from which all observed responses arise—is a linear time-invariant (LTI) system. The LTI approach assumes that the HDR is temporally and spatially stationary, permitting a single reference function to be used to test for the presence of "activation" in all acquired voxel time series, independent of the actual sequence of stimulus presentations to the subject. The assumptions of LTI systems have been documented to hold well in the visual system (Boynton et al., 1996), although extension to the auditory system is less clear (e.g., Talavage & Edmister, 2004; Olulade et al., 2011).

6.4.4.1 Basis Functions for Model-Based Analysis

Under the assumptions of LTI systems, the BOLD response to any stimulus or task can be modeled as the sum of multiple shifted and scaled instances of a single basis function (e.g., the HRF model of the HDR) or a small set of basis functions (e.g., the chosen HRF and derivatives). Convolution of the basis function(s) with a (binary) vector representing the presence and absence of the target condition thus provides a reference waveform depicting the expected signal levels to be acquired from the brain as a function of time due to the condition of interest. Through regression of this waveform against the observed signals, one obtains a measure of the strength (i.e., fit coefficient) of the response of every voxel in the brain to the particular target condition.

The HRF model typically used in fMRI analysis is based on a gamma variate function (see Fig. 6.1B for two examples). This model was first used to characterize the HDR in visual cortex (Boynton et al., 1996). As such, the onset, rise time, and duration that have been codified in most fMRI software may not accurately characterize responses elsewhere in the brain (Birn et al., 2001; Saad et al., 2003), including auditory cortex (Hu et al., 2010). If the true HDR differs in onset or duration from the reference function by even a few seconds, sensitivity could be severely affected (Lindquist et al., 2009), particularly when analyzing designs intended to detect or estimate transient signals (see Section 6.3.3).

One common means of correcting for variability in the HDR is regression using multiple reference waveform components, typically derived from a derivative-based expansion of an HRF kernel (Calhoun et al., 2004). Although this process enhances the fit to the observed time series, it reduces the degrees of freedom and hence the analytical power.

6.4.4.2 Assessment of Model-Based Statistical Significance

The assessment of the significance of a finding under model-based analysis is dependent on the goal of the analysis. Detection experiments are usually analyzed by testing the estimated effect size (e.g., difference in two fit coefficients, as obtained from a GLM analysis) against a null hypothesis of zero change. Estimation experiments are generally analyzed by characterization and comparison of the HRFs associated with each desired stimulus or task.

Under traditional Gaussian white-noise assumptions, the individual fit coefficients (i.e., parameter estimates; commonly referred to as "beta" values due to the symbolic coefficients in the model equation) obtained from GLM analysis using the chosen basis set may be converted into statistics with a known distribution (e.g., a *t*- or *z*-statistic). This processing results in a brain volume for each contrast (condition or comparison of conditions) in which the sensitivity of each voxel to the associated factor is indexed by a statistic.

6.4.4.3 Correcting for Multiple Comparisons

A typical fMRI brain volume contains several hundred thousand voxels, of which roughly 10% represent brain tissue. If the statistical threshold for each voxel is set to $p < 0.05$, then 5% of the voxels analyzed above (~10,000 for a whole-brain volume) will be erroneously observed to be "significant." To limit these false positives, it is typical to implement a correction for multiple comparisons before display of statistical maps. Bonferroni, familywise error (FWE), and false discovery rate (FDR) corrections are those most commonly applied in the literature. However, the former two approaches have been criticized for being overly conservative, whereas the latter may be interpreted as inappropriately liberal when there are a large number of positive tests (Bennett et al., 2009).

An alternative means to limit false positives is to identify clusters of contiguous voxels that, as an ensemble, are deemed unlikely to be observed as "active" by chance. The spatial smoothness of the data may be evaluated to estimate this chance level, permitting identification of the minimum cluster size required (commonly denoted *k* in the literature) to achieve a given test α-level, typically including a correction for multiple comparisons at the cluster level. Such cluster-based analyses are generally of greater sensitivity than single-voxel analyses followed by corrections for multiple comparisons, but there is a loss of precision regarding the location of activity, as the significance assessment applies to the entire cluster, rather than to any single voxel (and corresponding point on the brain) within the cluster.

Given the many ways in which significance can be computed, there is no "correct" threshold to use in a study, just as there is no "correct" way to analyze the data. In general, it may be assumed that the reported *p*-value threshold has been selected by the experimenter to be the strictest threshold that reveals what are *a priori* expected to be the key findings, yet still achieves an informal community threshold level (e.g., $p_{\mathrm{FWE}} < 0.05$, or $p_{\mathrm{FDR}} < 0.05$, or if none of these are successful, $p_{\mathrm{Uncorrected}} < 0.001$ is commonly accepted as "trend-level" significance). Although this variation is problematic in that errors are likely to exist in the published literature, appropriate explanation of the chosen thresholds allow the data to be of value to the community (Bennett et al., 2009). Further, it may be the false-negative (type II error) rate that should be minimized (Lieberman & Cunningham, 2009), and an unacknowledged desire to avoid false negatives (and a tolerance for false positives) may underlie the community's acceptance of variable thresholds.

Once a particular threshold is determined, *p*-values are generally presented as pseudocolor maps overlaid on grayscale images of high-resolution brain anatomy, either from an individual or an averaged image from a group. The locations of significant signal change can be identified using macroanatomical landmarks or anatomical databases (e.g., Brett et al., 2002; Shattuck et al., 2008) revealing the brain locations most sensitive to the effects of interest.

6.4.4.4 Region of Interest Analysis

Analysis is typically conducted at each voxel in the entire imaged volume, but a given experiment may have prior hypotheses regarding particular subregions of this volume. Region of interest (ROI) analysis is performed to limit analysis to just those regions.

ROIs can be defined in many ways, based on functional criteria, on observable anatomical structures, or on estimates of the underlying cytoarchitecture. Functional ROIs must be defined on an independent contrast, or in an independent group of subjects (see Kriegeskorte et al., 2009) but are readily generated by thresholding an activation image at a chosen level of statistical significance. A functional ROI can also be defined as a region of a particular size around a fixed coordinate taken from the published literature, provided the data are in the same anatomical reference space.

Anatomically defined ROIs are commonly based on observable anatomical landmarks (e.g., Heschl's gyrus); either from the anatomy of a single individual (e.g., Tzourio-Mazoyer et al., 2002), or probabilistically, based on anatomical variability in a group (e.g., Penhune et al., 1996; Shattuck et al., 2008). The ROI may also be defined from probabilistic maps of cytoarchitectonic regions, derived from maps of microanatomical features in several postmortem specimens (Eickhoff et al., 2005). In the realm of auditory neuroscience, several maps of auditory- and speech-related regions exist, including primary auditory cortex (e.g., Morosan et al., 2001) and inferior frontal cortex (Amunts et al., 1999).

Masks for anatomically defined regions as represented in a standard stereotactic space (Section 6.4.2), are commonly embedded in or compatible with neuroimaging data processing tools (e.g., Fischl et al., 2002; Eickhoff et al., 2005). These tools permit the researcher to focus an investigation on the subset of voxels most likely to correspond to a cortical region of interest. By means of these tools, one may evaluate condition-specific effects, either by averaging over voxels of a structure to compute mean percent signal change or a *t*-statistic, or by counting the number of voxels exceeding a given statistical height threshold. Note that ROIs are also often used to reduce the multiple-comparisons problem (Section 6.4.4.3)—when a smaller volume is assessed, a lesser correction is applied, enhancing the apparent detection power (Worsley et al., 1996). These "small-volume corrections" must be carefully applied because the risk of false positives increases with more liberal significance thresholds (Bennett et al., 2009).

6.4.5 Data-Driven Analyses

Whereas GLM approaches are model driven, other analyses seek to identify and characterize spatial patterns without reliance on fitting predetermined models.

6.4.5.1 Approaches for Data Reduction

Many data-driven analyses focus on reducing high-dimensional data into a smaller set of components that account for a large portion of the variation in the data. These factor analytic approaches include principal components analysis, independent components analysis, and partial least squares analysis (McIntosh et al., 1996), which is a hybrid model- and data-driven approach. Unlike model-driven approaches, in which a model is typically fit separately to each voxel, these approaches tend to simultaneously consider all voxels in an imaging volume.

Independent components analysis (ICA) is a popular data-driven technique that extracts, from the time-series data, a set of spatiotemporal patterns that are maximally different from each other. It is used to identify sets of brain regions that exhibit similar time courses over the duration of an experiment, with this similarity interpreted as being indicative of common underlying function, particularly interconnectedness through a specific processing network (e.g., Schmithorst & Holland, 2004b). Note that this approach lends itself well both to analysis of stimulus- or task-driven responses (i.e., where the experimenter imposes a structure based on the relationship among conditions) and to studies of connectivity (see Section 6.4.6). ICA and other data reduction techniques are extremely powerful in that common activity may be identified in the absence of an assumption about the form of the HDR, although evaluation of the functional relevance of a particular component (i.e., whether it embodies the response to an effect of interest) often requires falling back on the GLM approaches described earlier. One limitation of ICA is that it is sensitive to low-frequency physiological fluctuations (Birn et al., 2008). Another is that there is no absolute standard for selection of the number of relevant components, and this is an extremely important parameter that can dramatically affect results; but see probabilistic ICA (Beckmann & Smith, 2004) for a less arbitrary process.

6.4.5.2 Pattern Classification

Pattern classification algorithms are another multivariate approach to data analysis and are used to study patterns of activity across voxels and determine whether, and how, these patterns change as a function of stimulus features. Also known as multivoxel pattern analysis (MVPA), this approach exploits the multivariate nature of brain imaging data to determine how mental representations (of stimulus features or categories) map on distributed patterns of neural activity (Kamitani & Tong, 2005). This approach, also referred to as "decoding" or even "brain reading" (Cox & Savoy, 2003), represents a new application of statistical pattern recognition.

Pattern classification algorithms are predicated on the observation that, if the experimental conditions (that give rise to the mental representations of interest) can be classified at better-than-chance levels solely on the basis of the distribution of activity in the brain, then this activity pattern must carry appreciable information about the experimental conditions. Excitingly, this method has been used to reveal the dimensions that the brain uses to categorize stimuli, even in the absence of predefined hypotheses (Kriegeskorte, 2011) or paradigms for presentation of experimental conditions (Polyn et al., 2005).

6.4.6 Functional and Effective Connectivity

The simplest form of connectivity analysis assesses correlations between time series from two brain regions. This will reveal, among other things, coactivation—that is, the joint sensitivity of both regions to an experimental effect of interest—which will also be revealed by conventional analysis. In the case of task-based assessments of connectivity, observation of the coupling in the moment-to-moment fluctuations in activity between two regions generally requires that the variability in the time series due to experimental factors must first be accounted for, typically through traditional GLM analysis (Horwitz, 2003).

It is important to recognize that functional connectivity analysis does not identify the direction of control that regions have over one another, and Friston et al. (1993) proposed that "effective connectivity" be defined as the influence that one neural system exerts over another in a directional, causal sense. Several alternative analytic approaches have been used to characterize such connections. Assessment of "effective connectivity" has been performed with a variety of approaches (e.g., Buchel & Friston, 1997; Friston et al., 2003; Goebel et al., 2003) and used in auditory fMRI to address questions concerning the strength of connections among auditory cortical regions (Harrison et al., 2003).

6.5 Summary

Hemodynamic imaging methods, despite only indirectly reflecting neural activity, remain powerful tools for the study of the auditory system. The potential of fMRI as a tool to study functional organization and behavior in a longitudinal manner provides rich opportunities for both basic and applied human research in hearing loss; compensation for hearing loss; and development of auditory functions and other changes with age, experience, and physiological state. Further, advances in nonhuman primate imaging open the door to comparative studies that will be invaluable sources of information, given that so much more is currently known about the organization of, for example, the macaque auditory system as compared to that of the human auditory system. However, techniques are presently being developed that may meaningfully reduce the confounds associated with the imaging acoustic noise in an experimental and/or post-processing setting, and other methodological developments to study connectivity and to conduct pattern classification analysis increase the utility of fMRI. Accordingly, the importance of fMRI as a research tool for the auditory neuroscience community will probably continue to increase, expanding knowledge of sensory and cognitive processes both in diseased and healthy states.

Acknowledgments ISJ is supported by the Canada Research Chairs Program and the NSERC E. W. R. Steacie Memorial Fellowship Program. TMT and JGC gratefully acknowledge the support and assistance of our colleagues, Drs. Gavin Bidelmann, Alexander L. Francis, Shuowen Hu, David L. Kemmerer, Olumide Olulade, Ruwan D. Ranaweera and Gregory G. Tamer, Jr., and the invaluable input and contributions from Christopher Smalt.

References

Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). The variability of human, BOLD hemodynamic responses. *NeuroImage*, 8(4), 360–369.

Amunts, K., Schleicher, A., Burgel, U., Mohlberg, H., Uylings, H. B., & Zilles, K. (1999). Broca's region revisited: Cytoarchitecture and intersubject variability. *Journal of Comparative Neurology*, 412(2), 319–341.

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage*, 38(1), 95–113.

Aubert, A., & Costalat, R. (2002). A model of the coupling between brain electrical activity, metabolism, and hemodynamics: Application to the interpretation of functional neuroimaging. *NeuroImage*, 17(3), 1162–1181.

Baciu, M., Watson, J., Maccotta, L., McDermott, K., Buckner, R., Gilliam, F., & Ojemann, J. (2005). Evaluating functional MRI procedures for assessing hemispheric language dominance in neurosurgical patients. *Neuroradiology*, 47, 835–844.

Bandettini, P. A., & Cox, R. W. (2000). Event-related fMRI contrast when using constant interstimulus interval: Theory and experiment. *Magnetic Resonance in Medicine*, 43(4), 540–548.

Bandettini, P. A., Jesmanowicz, A., Van Kylen, J., Birn, R. M., & Hyde, J. S. (1998). Functional MRI of brain activation induced by scanner acoustic noise. *Magnetic Resonance in Medicine*, 39(3), 410–416.

Beckmann, C. F., & Smith, S. M. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, 23(2), 137–152.

Belin, P., Zatorre, R. J., Hoge, R., Evans, A. C., & Pike, B. (1999). Event-related fMRI of the auditory cortex. *NeuroImage*, 10(4), 417–429.

Belliveau, J. W., Rosen, B. R., Kantor, H. L., Rzedzian, R. R., Kennedy, D. N., McKinstry, R. C., et al. (1990). Functional cerebral imaging by susceptibility-contrast NMR. *Magnetic Resonance in Medicine*, 14(3), 538–546.

Belliveau, J. W., Kennedy, D. N., Jr., McKinstry, R. C., Buchbinder, B. R., Weisskoff, R. M., Cohen, M. S., et al. (1991). Functional mapping of the human visual cortex by magnetic resonance imaging. *Science*, 254(5032), 716–719.

Bennett, C. M., Wolford, G. L., & Miller, M. B. (2009). The principled control of false positives in neuroimaging. *Social Cognitive and Affective Neuroscience*, 4(4), 417–422.

Binder, J. R., Rao, S. M., Hammeke, T. A., Frost, J. A., Bandettini, P. A., & Hyde, J. S. (1994). Effects of stimulus rate on signal response during functional magnetic-resonance-imaging of auditory-cortex. *Cognitive Brain Research*, 2(1), 31–38.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Rao, S. M., & Cox, R. W. (1999). Conceptual processing during the conscious resting state: A functional MRI study. *Journal of Cognitive Neuroscience*, 11(1), 80–93.

Birn, R. M., Saad, Z. S., & Bandettini, P. A. (2001). Spatial heterogeneity of the nonlinear dynamics in the FMRI BOLD response. *NeuroImage*, 14(4), 817–826.

Birn, R. M., Cox, R. W., & Bandettini, P. A. (2002). Detection versus estimation in event-related fMRI: Choosing the optimal stimulus timing. *NeuroImage*, 15(1), 252–264.

Birn, R. M., Murphy, K., & Bandettini, P. A. (2008). The effect of respiration variations on independent component analysis results of resting state functional connectivity. *Human Brain Mapping*, 29(7), 740–750.

Boxerman, J. L., Bandettini, P. A., Kwong, K. K., Baker, J. R., Davis, T. L., Rosen, B. R., & Weisskoff, R. M. (1995). The intravascular contribution to fMRI signal change: Monte Carlo modeling and diffusion-weighted studies in vivo. *Magnetic Resonance in Medicine*, 34(1), 4–10.

Boynton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *Journal of Neuroscience*, 16(13), 4207–4221.

Brett, M., Johnsrude, I. S., & Owen, A. M. (2002). The problem of functional localization in the human brain. *Nature Reviews Neuroscience*, 3(3), 243–249.

Büchel, C., & Friston, K. J. (1997). Modulation of connectivity in visual pathways by attention: Cortical interactions evaluated with structural equation modelling and fMRI. *Cerebral Cortex*, 7(8), 768–778.

Büchel, C., Holmes, A. P., Rees, G., & Friston, K. J. (1998). Characterizing stimulus-response functions using nonlinear regressors in parametric fMRI experiments. *NeuroImage*, 8(2), 140–148.

Buckner, R. L. (1998). Event-related fMRI and the hemodynamic response. *Human Brain Mapping*, 6(5–6), 373–377.

Burock, M. A., Buckner, R. L., Woldorff, M. G., Rosen, B. R., & Dale, A. M. (1998). Randomized event-related experimental designs allow for extremely rapid presentation rates using functional MRI. *NeuroReport*, 9(16), 3735–3739.

Calhoun, V. D., Adali, T., Pearlson, G. D., & Pekar, J. J. (2001). A method for making group inferences from functional MRI data using independent component analysis. *Human Brain Mapping*, 14(3), 140–151.

Calhoun, V. D., Stevens, M. C., Pearlson, G. D., & Kiehl, K. A. (2004). fMRI analysis with the general linear model: Removal of latency-induced amplitude bias by incorporation of hemodynamic derivative terms. *NeuroImage*, 22(1), 252–257.

Chakraborty, A., & McEvoy, A. W. (2008). Presurgical functional mapping with functional MRI. *Current Opinion in Neurology*, 21(4), 446–451.

Cox, R. W., & Jesmanowicz, A. (1999). Real-time 3D image registration for functional MRI. *Magnetic Resonance in Medicine*, 42(6), 1014–1018.

Cox, D. D., & Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) "brain reading": Detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage*, 19(2 Pt 1), 261–270.

Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *Journal of Neuroscience*, 23(8), 3423–3431.

Disbrow, E. A., Slutsky, D. A., Roberts, T. P. L., & Krubitzer, L. A. (2000). Functional MRI at 1.5 tesla: A comparison of the blood oxygenation level-dependent signal and electrophysiology. *Proceedings of the National Academy of Sciences of the USA*, 97(17), 9718–9723.

Eden, G. F., Joseph, J. E., Brown, H. E., Brown, C. P., & Zeffiro, T. A. (1999). Utilizing hemodynamic delay and dispersion to detect fMRI signal change without auditory interference: The behavior interleaved gradients technique. *Magnetic Resonance in Medicine*, 41(1), 13–20.

Edmister, W. B., Talavage, T. M., Ledden, P. J., & Weisskoff, R. M. (1999). Improved auditory cortex imaging using clustered volume acquisitions. *Human Brain Mapping*, 7(2), 89–97.

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage*, 25(4), 1325–1335.

Fischl, B., Sereno, M. I., Tootell, R. B., & Dale, A. M. (1999). High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping*, 8(4), 272–284.

Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., et al. (2002). Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3), 341–355.

Fischl, B., Rajendran, N., Busa, E., Augustinack, J., Hinds, O., Yeo, B. T., et al. (2008). Cortical folding patterns and predicting cytoarchitecture. *Cerebral Cortex*, 18(8), 1973–1980.

Foster, J. R., Hall, D. A., Summerfield, A. Q., Palmer, A. R., & Bowtell, R. W. (2000). Sound-level measurements and calculations of safe noise dosage during EPI at 3 T. *Journal of Magnetic Resonance Imaging*, 12(1), 157–163.

Frank, L. R., Buxton, R. B., & Wong, E. C. (2001). Estimation of respiration-induced noise fluctuations from undersampled multislice fMRI data. *Magnetic Resonance in Medicine*, 45(4), 635–644.

Friston, K. J., Frith, C., & Frackowiak, R. (1993). Time-dependent changes in effective connectivity measured with PET. *Human Brain Mapping*, 1, 69–79.

Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C., Frackowiak, R. S., & Turner, R. (1995). Analysis of fMRI time-series revisited. *NeuroImage*, 2(1), 45–53.

Friston, K. J., Price, C. J., Fletcher, P., Moore, C., Frackowiak, R. S., & Dolan, R. J. (1996). The trouble with cognitive subtraction. *NeuroImage*, 4(2), 97–104.

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, 19(4), 1273–1302.

Gati, J. S., Menon, R. S., Ugurbil, K., & Rutt, B. K. (1997). Experimental determination of the BOLD field strength dependence in vessels and tissue. *Magnetic Resonance in Medicine*, 38(2), 296–302.

Giraud, A. L., Lorenzi, C., Ashburner, J., Wable, J., Johnsrude, I., Frackowiak, R., & Kleinschmidt, A. (2000). Representation of the temporal envelope of sounds in the human brain. *Journal of Neurophysiology*, 84(3), 1588–1598.

Gjedde, A., & Marrett, S. (2001). Glycolysis in neurons, not astrocytes, delays oxidative metabolism of human visual cortex during sustained checkerboard stimulation in vivo. *Journal of Cerebral Blood Flow and Metabolism*, 21(12), 1384–1392.

Glover, G. H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage*, 9(4), 416–429.

Glover, G. H., Lemieux, S. K., Drangova, M., & Pauly, J. M. (1996). Decomposition of inflow and blood oxygen level-dependent (BOLD) effects with dual-echo spiral gradient-recalled echo (GRE) fMRI. *Magnetic Resonance in Medicine*, 35(3), 299–308.

Glover, G. H., Li, T. Q., & Ress, D. (2000). Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. *Magnetic Resonance in Medicine*, 44(1), 162–167.

Goebel, R., Roebroeck, A., Kim, D. S., & Formisano, E. (2003). Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. *Magnetic Resonance Imaging*, 21(10), 1251–1261.

Goense, J. B., & Logothetis, N. K. (2008). Neurophysiology of the BOLD fMRI signal in awake monkeys. *Current Biology*, 18(9), 631–640.

Grady, C. L., Van Meter, J. W., Maisog, J. M., Pietrini, P., Krasuski, J., & Rauschecker, J. P. (1997). Attention-related modulation of activity in primary and secondary auditory cortex. *NeuroReport*, 8(11), 2511–2516.

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, 48(1), 63–72.

Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., et al. (1999). "Sparse" temporal sampling in auditory fMRI. *Human Brain Mapping*, 7(3), 213–223.

Hall, D. A., Chambers, J., Akeroyd, M. A., Foster, J. R., Coxon, R., & Palmer, A. R. (2009). Acoustic, psychophysical, and neuroimaging measurements of the effectiveness of active cancellation during auditory functional magnetic resonance imaging. *Journal of the Acoustical Society of America*, 125(1), 347–359.

Handwerker, D. A., Ollinger, J. M., & D'Esposito, M. (2004). Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *NeuroImage*, 21(4), 1639–1651.

Harms, M. P., & Melcher, J. R. (2002). Sound repetition rate in the human auditory pathway: Representations in the waveshape and amplitude of fMRI activation. *Journal of Neurophysiology*, 88(3), 1433–1450.

Harms, M. P., & Melcher, J. R. (2003). Detection and quantification of a wide range of fMRI temporal responses using a physiologically-motivated basis set. [Research Support, Non-U.S. Gov't]. *Human Brain Mapping*, 20(3), 168–183.

Harrison, L., Penny, W. D., & Friston, K. (2003). Multivariate autoregressive modeling of fMRI time series. *NeuroImage*, 19(4), 1477–1491.

Hennig, J., Ernst, T., Speck, O., Deuschl, G., & Feifel, E. (1994). Detection of brain activation using oxygenation sensitive functional spectroscopy. *Magnetic Resonance in Medicine*, 31(1), 85–90.

Henson, R. (2006a). Forward inference using functional neuroimaging: Dissociations versus associations. *Trends in Cognitive Sciences*, 10(2), 64–69.

Henson, R. (2006b). Efficient experimental design for fMRI. In K. Friston, J. Ashburner, S. Kiebel, T. Nichols, & W. Penny (Eds.), *Statistical parametric mapping: The analysis of functional brain images* (pp. 193–210). London: Academic Press.

Henson, R., Buchel, C., Josephs, O., & Fristen, K. (1999). The slice-timing problem in event-related fMRI. *NeuroImage*, 9(6 Part II).

Herdener, M., Esposito, F., di Salle, F., Boller, C., Hilti, C. C., Habermeyer, B., et al. (2010). Musical training induces functional plasticity in human hippocampus. *Journal of Neuroscience*, 30(4), 1377–1384.

Horwitz, B. (2003). The elusive concept of brain connectivity. *NeuroImage*, 19(2 Pt 1), 466–470.

Hu, S., Olulade, O., Castillo, J. G., Santos, J., Kim, S., Tamer, G. G., Jr., et al. (2010). Modeling hemodynamic responses in auditory cortex at 1.5 T using variable duration imaging acoustic noise. *NeuroImage*, 49(4), 3027–3038.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, 8(5), 679–685.

Kriegeskorte, N. (2011). Pattern-information analysis: From stimulus decoding to computational-model testing. *NeuroImage*, 56(2), 411–421.

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience*, 12(5), 535–540.

Krings, T., Schreckenberger, M., Rohde, V., Foltys, H., Spetzger, U., Sabri, O., et al. (2001). Metabolic and electrophysiological validation of functional MRI. *Journal of Neurology, Neurosurgery and Psychiatry*, 71(6), 762–771.

Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., et al. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences of the USA*, 89(12), 5675–5679.

Lauterbur, P. C. (1973). Image formation by induced local interactions: Examples employing nuclear magnetic resonance. *Nature*, 242, 190–191.

Le, T. H., Patel, S., & Roberts, T. P. (2001). Functional MRI of human auditory cortex using block and event-related designs. *Magnetic Resonance in Medicine*, 45(2), 254–260.

Lieberman, M. D., & Cunningham, W. A. (2009). Type I and type II error concerns in fMRI research: Re-balancing the scale. *Social Cognitive and Affective Neuroscience*, 4(4), 423–428.

Lin, F. H., Kwong, K. K., Belliveau, J. W., & Wald, L. L. (2004). Parallel imaging reconstruction using automatic regularization. *Magnetic Resonance in Medicine*, 51(3), 559–567.

Lindquist, M. A., Meng Loh, J., Atlas, L. Y., & Wager, T. D. (2009). Modeling the hemodynamic response function in fMRI: Efficiency, bias and mis-modeling. *NeuroImage*, 45(1 Supplement), S187–198.

Loenneker, T., Hennel, F., Ludwig, U., & Hennig, J. (2001). Silent BOLD imaging. *Magnetic Resonance Materials in Physics Biology and Medicine*, 13(2), 76–81.

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, 453(7197), 869–878.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412(6843), 150–157.

Lowe, M. J., Mock, B. J., & Sorenson, J. A. (1998). Functional connectivity in single and multislice echoplanar imaging using resting-state fluctuations. *NeuroImage*, 7(2), 119–132.

Malonek, D., & Grinvald, A. (1996). Interactions between electrical activity and cortical microcirculation revealed by imaging spectroscopy: Implications for functional brain mapping. *Science*, 272(5261), 551–554.

Malonek, D., Dirnagl, U., Lindauer, U., Yamada, K., Kanno, I., & Grinvald, A. (1997). Vascular imprints of neuronal activity: Relationships between the dynamics of cortical blood flow, oxygenation, and volume changes following sensory stimulation. *Proceedings of the National Academy of Sciences of the USA*, 94(26), 14826–14831.

Mandeville, J. B., Marota, J. J., Kosofsky, B. E., Keltner, J. R., Weissleder, R., Rosen, B. R., & Weisskoff, R. M. (1998). Dynamic functional imaging of relative cerebral blood volume during rat forepaw stimulation. *Magnetic Resonance in Medicine*, 39(4), 615–624.

Mansfield, P. (1977). Multi-planar image formation using NMR spin echoes. *Journal of Physics C: Solid State Physics*, 10, L55–L58.

McIntosh, A. R., Bookstein, F. L., Haxby, J. V., & Grady, C. L. (1996). Spatial pattern analysis of functional brain images using partial least squares. *NeuroImage*, 3(3 Pt 1), 143–157.

Mechefske, C. K., Geris, R., Gati, J. S., & Rutt, B. K. (2001). Acoustic noise reduction in a 4 T MRI scanner. *Magnetic Resonance Materials in Physics*, *Biology and Medicine*, 13(3), 172–176.

Menon, R. S., & Kim, S. G. (1999). Spatial and temporal limits in cognitive neuroimaging with fMRI. *Trends in Cognitive Sciences*, 3(6), 207–216.

Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T., & Zilles, K. (2001). Human primary auditory cortex: Cytoarchitectonic subdivisions and mapping into a spatial reference system. *NeuroImage*, 13(4), 684–701.

Norris, D. G. (2006). Principles of magnetic resonance assessment of brain function. *Journal of Magnetic Resonance Imaging*, 23(6), 794–807.

Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the USA*, 87(24), 9868–9872.

Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S. G., Merkle, H., & Ugurbil, K. (1992). Intrinsic signal changes accompanying sensory stimulation: Functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Sciences of the USA*, 89(13), 5951–5955.

Ojemann, J. G., Akbudak, E., Snyder, A. Z., McKinstry, R. C., Raichle, M. E., & Conturo, T. E. (1997). Anatomic localization and quantitative analysis of gradient refocused echo-planar fMRI susceptibility artifacts. *NeuroImage*, 6(3), 156–167.

Olulade, O., Hu, S., Gonzalez-Castillo, J., Tamer, G. G., Jr., Luh, W. M., Ulmer, J. L., & Talavage, T. M. (2011). Assessment of temporal state-dependent interactions between auditory fMRI responses to desired and undesired acoustic sources. *Hearing Research*, 277(1–2), 67–77.

Orfanidou, E., Marslen-Wilson, W. D., & Davis, M. H. (2006). Neural response suppression predicts repetition priming of spoken words and pseudowords. *Journal of Cognitive Neuroscience*, 18(8), 1237–1252.

Peelle, J. E., Eason, R. J., Schmitter, S., Schwarzbauer, C., & Davis, M. H. (2010). Evaluating an acoustically quiet EPI sequence for use in fMRI studies of speech and auditory processing. *NeuroImage*, 52(4), 1410–1419.

Penhune, V. B., Zatorre, R. J., MacDonald, J. D., & Evans, A. C. (1996). Interhemispheric anatomical differences in human primary auditory cortex: Probabilistic mapping and volume measurement from magnetic resonance scans. *Cerebral Cortex*, 6(5), 661–672.

Poldrack, R. A., & Wagner, A. D. (2004). What can neuroimaging tell us about the mind? Insights from prefrontal cortex. *Current Directions in Psychological Science*, 13(5), 177–181.

Polyn, S. M., Natu, V. S., Cohen, J. D., & Norman, K. A. (2005). Category-specific cortical activity precedes retrieval during memory search. *Science*, 310(5756), 1963–1966.

Pruessmann, K. P., Weiger, M., Scheidegger, M. B., & Boesiger, P. (1999). SENSE: Sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine*, 42(5), 952–962.

Rademacher, J., Caviness, V. S., Jr., Steinmetz, H., & Galaburda, A. M. (1993). Topographical variation of the human primary cortices: Implications for neuroimaging, brain mapping, and neurobiology. *Cerebral Cortex*, 3(4), 313–329.

Raichle, M. E. (2009). A brief history of human brain mapping. *Trends in Neurosciences*, 32(2), 118–126.

Ravicz, M. E., & Melcher, J. R. (2001). Isolating the auditory system from acoustic noise during functional magnetic resonance imaging: Examination of noise conduction through the ear canal, head, and body. *Journal of the Acoustical Society of America*, 109(1), 216–231.

Ravicz, M. E., Melcher, J. R., & Kiang, N. Y. (2000). Acoustic noise during functional magnetic resonance imaging. *Journal of the Acoustical Society of America*, 108(4), 1683–1696.

Roy, C. S., & Sherrington, C. S. (1890). On the regulation of the blood-supply of the brain. *Journal of Physiology*, 11(1–2), 85–108.

Saad, Z. S., Ropella, K. M., Cox, R. W., & DeYoe, E. A. (2001). Analysis and use of FMRI response delays. *Human Brain Mapping*, 13(2), 74–93.

Saad, Z. S., DeYoe, E. A., & Ropella, K. M. (2003). Estimation of FMRI response delays. *NeuroImage*, 18(2), 494–504.

Sammler, D., Baird, A., Valabregue, R., Clement, S., Dupont, S., Belin, P., & Samson, S. (2010). The relationship of lyrics and tunes in the processing of unfamiliar songs: A functional magnetic resonance adaptation study. *Journal of Neuroscience*, 30(10), 3572–3578.

Schmithorst, V. J., & Holland, S. K. (2004a). Event-related fMRI technique for auditory processing with hemodynamics unrelated to acoustic gradient noise. *Magnetic Resonance in Medicine*, 51(2), 399–402.

Schmithorst, V. J., & Holland, S. K. (2004b). Comparison of three methods for generating group statistical inferences from independent component analysis of functional magnetic resonance imaging data. *Journal of Magnetic Resonance Imaging*, 19(3), 365–368.

Schonwiesner, M., & Zatorre, R. J. (2009). Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fMRI. *Proceedings of the National Academy of Sciences of the USA*, 106(34), 14611–14616.

Schwarzbauer, C., Davis, M. H., Rodd, J. M., & Johnsrude, I. (2006). Interleaved silent steady state (ISSS) imaging: A new sparse imaging method applied to auditory fMRI. *NeuroImage*, 29(3), 774–782.

Seifritz, E., Di Salle, F., Esposito, F., Herdener, M., Neuhoff, J. G., & Scheffler, K. (2006). Enhancing BOLD response in the auditory system by neurophysiologically tuned fMRI sequence. *NeuroImage*, 29(3), 1013–1022.

Shattuck, D. W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K. L., et al. (2008). Construction of a 3D probabilistic atlas of human cortical structures. *NeuroImage*, 39(3), 1064–1080.

Sheth, S. A., Nemoto, M., Guiou, M. W., Walker, M. A., & Toga, A. W. (2005). Spatiotemporal evolution of functional hemodynamic changes and their relationship to neuronal activity. *Journal of Cerebral Blood Flow and Metabolism*, 25(7), 830–841.

Shmuel, A., Yacoub, E., Pfeuffer, J., Van de Moortele, P. F., Adriany, G., Hu, X., & Ugurbil, K. (2002). Sustained negative BOLD, blood flow and oxygen consumption response and its coupling to the positive response in the human brain. *Neuron*, 36(6), 1195–1210.

Shmuel, A., Augath, M., Oeltermann, A., & Logothetis, N. K. (2006). Negative functional MRI response correlates with decreases in neuronal activity in monkey visual area V1. *Nature Neuroscience*, 9(4), 569–577.

Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain:* 3-*Dimensional proportional system—an approach to cerebral imaging*. New York: Thieme.

Talavage, T. M., & Edmister, W. B. (2004). Nonlinearity of FMRI responses in human auditory cortex. *Human Brain Mapping*, 22(3), 216–228.

Talavage, T. M., Ledden, P. J., Benson, R. R., Rosen, B. R., & Melcher, J. R. (2000). Frequency-dependent responses exhibited by multiple regions in human auditory cortex. *Hearing Research*, 150(1–2), 225–244.

Tomasi, D., Caparelli, E. C., Chang, L., & Ernst, T. (2005). fMRI-acoustic noise alters brain activation during working memory tasks. *NeuroImage*, 27(2), 377–386.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, 15(1), 273–289.

Uludag, K., Muller-Bierl, B., & Ugurbil, K. (2009). An integrative model for neuronal activity-induced signal changes for gradient and spin echo functional imaging. *NeuroImage*, 48(1), 150–165.

Viswanathan, A., & Freeman, R. D. (2007). Neurometabolic coupling in cerebral cortex reflects synaptic more than spiking activity. *Nature Neuroscience*, 10(10), 1308–1312.

Weigelt, S., Muckli, L., & Kohler, A. (2008). Functional magnetic resonance adaptation in visual neuroscience. *Reviews in the Neurosciences*, 19(4–5), 363–380.

Worsley, K. J., & Friston, K. J. (1995). Analysis of fMRI time-series revisited–again. *NeuroImage*, 2(3), 173–181.

Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., & Evans, A. C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping*, 4(1), 58–73.

Wowk, B., McIntyre, M. C., & Saunders, J. K. (1997). k-Space detection and correction of physiological artifacts in fMRI. *Magnetic Resonance in Medicine*, 38(6), 1029–1034.

Zhang, N., Zhu, X. H., & Chen, W. (2005). Influence of gradient acoustic noise on fMRI response in the human visual cortex. *Magnetic Resonance in Medicine*, 54(2), 258–263.

Part II
The Principal Computational Challenges

Chapter 7
Coding of Basic Acoustical and Perceptual Components of Sound in Human Auditory Cortex

Deborah Hall and Daphne Barker

7.1 Introduction

Neuroimaging studies are important for developing an understanding of the functional organization of human auditory cortex. This chapter summarizes the contributions from human neuroimaging studies that have examined cortical responses to a range of different sound stimuli. Although somewhat simpler than natural sounds, laboratory-generated sounds represent fundamental elements that are nonetheless important because they enable tight experimental control over other potentially confounding acoustical variables such as irregular variations in spectral complexity, spatial position, and level over time. Synthesized sound elements of interest include single-frequency and broadband spectra, sound level, sinusoidal spectrotemporal modulation, and pitch. Experimental studies that search for the cortical representation of these sound features are mostly presented from the field of functional magnetic resonance imaging (fMRI) (Talavage, Johnsrude, and Gonzalez Castillo, Chapter 6), but findings from other neuroimaging modalities are also reported. The chapter concludes (Section 7.7) with some examples of how novel approaches to experimental design and analysis are beginning to reveal how auditory stimulus attributes have spatially overlapping organizations.

D. Hall (✉)
NIHR National Biomedical Research Unit in Hearing, Nottingham NG1 5DU, UK
e-mail: deb.hall@nottingham.ac.uk

D. Barker
School of Psychological Sciences, University of Manchester, Manchester M13 9PL, UK
e-mail: daphne.barker@manchester.ac.uk

D. Poeppel et al. (eds.), *The Human Auditory Cortex*, Springer Handbook of Auditory Research 43, DOI 10.1007/978-1-4614-2314-0_7,

7.1.1 A Scheme for Parcellating Human Auditory Cortex

Most neuroimaging work on the human brain has focused on the functional architecture of macroscopic brain areas. This focus has been largely influenced by available methodology. For example, most experimental designs use time-integrated averaging procedures and usually analyze the data by means of subtracting one stimulus condition from another. fMRI acquisition protocols often use a 3 mm^3 resolution, and the data from neighboring volume elements are averaged (via spatial smoothing) to reduce noise (Talavage, Johnsrude, and Gonzalez Castillo, Chapter 6).

To interpret the areas of feature-sensitive activation with reference to the underlying neuroanatomy, auditory neuroscientists have made widespread use of supplementary information obtained using anatomical mapping techniques and functional recording methods in animals and in humans. In the case of noninvasive recordings of human central auditory function using neuroimaging methods such as fMRI, there is no definitive approach for parcellating living human auditory cortex into its major microanatomical divisions. A traditional strategy in the neurosciences has been to link specific auditory processes to their gyral and sulcal locations in the human brain because it has been understood that these macroscopic anatomical landmarks had an important physiological relevance. However, the advent of more sophisticated methods for studying the microanatomy has shown this to be a rather simplistic view of structure–function relations. Today, the neuroimaging field relies heavily on the results of electrophysiological and anatomical studies in animals and on postmortem studies of human anatomy to interpret and to localize human functional data. In this chapter, both macroanatomical and microanatomical approaches are discussed and the specific terminology used in this chapter is introduced here. Clarke and Morosan (Chapter 2) provide a more detailed description of the different schemes for parcellating human auditory cortex.

The key macroscopic features defining Heschl's gyrus, planum temporale, and planum polare (Fig. 7.1A) are consistently present, and these major macroanatomical landmarks are visible in vivo in MR scans of the human brain. Sound-related activity usually covers parts of these three regions, and this is especially true for hearing acoustically complex sounds and for tasks that involve active listening. The concepts of "core," "belt," and "parabelt" regions provide the basis for the organization of auditory cortex across numerous primate species, including humans (review: Hackett, 2003; see Fig. 7.1A). In humans, the core is typically centered on the medial two-thirds of Heschl's gyrus (Figs. 7.1B, C). In both primates and humans,

Fig. 7.1 (**a**) Surface of human left hemisphere with a cut through the Sylvian fissure to reveal the macroanatomical structure of auditory cortex on the inner surface, including Heschl's gyrus, planum polare and planum temporale. In this panel, the position of Heschl's gyrus (the core region) is shown by the dotted grey region. A suggestion for how belt and parabelt regions might be organized is shown by the dark (belt) and light grey (parabelt) shading. (**b**) Summary diagram of the

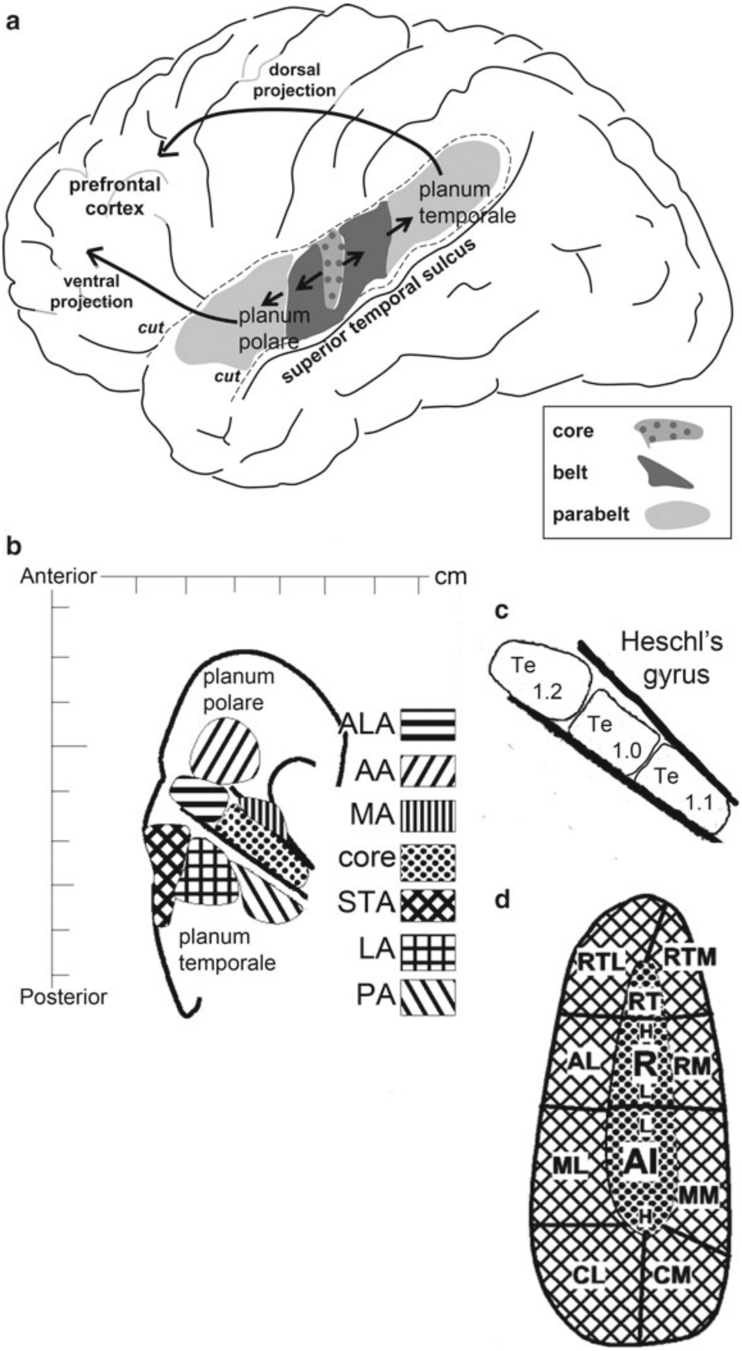


Fig. 7.1 (continued) microanatomical structure of the human supratemporal plane (left hemisphere) based on modifications of Figure 10 in Rivier and Clarke (1997) and Figure 6 in Wallace et al. (2002). (**c**) A closer look at Heschl's gyrus illustrates the microanatomical structure adopted in Chapter 7 (c.f. Morosan et al., 2001). (**d**) Summary diagram of the microanatomical structure of auditory cortex in macaque monkey (Kaas & Hackett, 2000). In both **b** and **d**, regions corresponding to the auditory core are dotted and regions possibly corresponding to the auditory belt are hatched. See text for an explanation of the abbreviations

the core appears to contain two fields: A1 and R in primates (Fig. 7.1D) and Te 1.0 and Te 1.1 in humans (Fig. 7.1C; Morosan et al., 2001). Very little is known about the adjacent field (RT in primates), but the parcellation scheme of Morosan et al. (2001) suggests that the human homologue may be area Te 1.2 on the lateral portion of Heschl's gyrus (Fig. 7.1C).

One of the goals that still motivates many human neuroimaging studies concerns the relationship between the localization of functional activity and the underlying microanatomy. Where it is possible to do so, the cortical representation of basic acoustic constituents are interpreted in terms of both macroanatomical and microanatomical definitions. In this chapter, Section 7.2.1 draws heavily on the delineation of Heschl's gyrus into Te 1.0, 1.1, and 1.2 when describing the pattern of frequency-dependent responses that characterize the tonotopic organization of human primary auditory cortex. The location of activity spanning the belt and parabelt regions draws upon the scheme defined by Rivier and Clarke (1997) and later confirmed by Wallace et al. (2002) (see Fig. 7.1B). Hence, Sections 7.2.2, 7.4.2, 7.4.3, and 7.5.2 all consider the localization of functional responses within the five nonprimary regions identified beyond Heschl's gyrus. The schemes of Morosan et al. (2001) and Rivier and Clarke (1997) are popular for speculating on the underlying microanatomical landscape of the observed feature-related auditory activity. Perhaps one of the main reasons for their favor is attributable to the authors' efforts to present their schemes in formats that are compatible with human functional images, most notably in terms of their transformation into a brain space that has standardized three-dimensional coordinates.

7.2 Single-Frequency Tones

Single-frequency tones (sinusoids) are the simplest type of acoustic signal because they form the building blocks from which all natural sounds can be expressed. Indeed, such form of frequency segregation is naturally performed by the cochlea for frequencies ranging from 20 Hz to 20 kHz. When a sinusoidal sound pressure wave is transmitted to the inner ear, it maximally vibrates a single place along the basilar membrane that is frequency specific (Fig. 7.2A). Hair cells at the place of maximum vibration serve to transduce the mechanical energy into neural impulses. Hence, taken along its entire length, the basilar membrane can be thought of as behaving like a series of frequency channels transmitting frequency information to the auditory nerve (Fig. 7.2B). In reality, the amount of excitation along the basilar membrane is not discrete but rather it decreases with successive shifts away from the best frequency. The resultant neural tuning curve reflects the degree of frequency selectivity (or width of each frequency channel). Using psychophysical methods, the width of a frequency channel has been estimated to be about 12% of the center frequency, for frequencies between 750 Hz and 5 kHz (Moore, 2004).

The gradient of frequency-specific coding along the cochlea is known as cochleotopy, although this orderly representation is maintained throughout the ascending

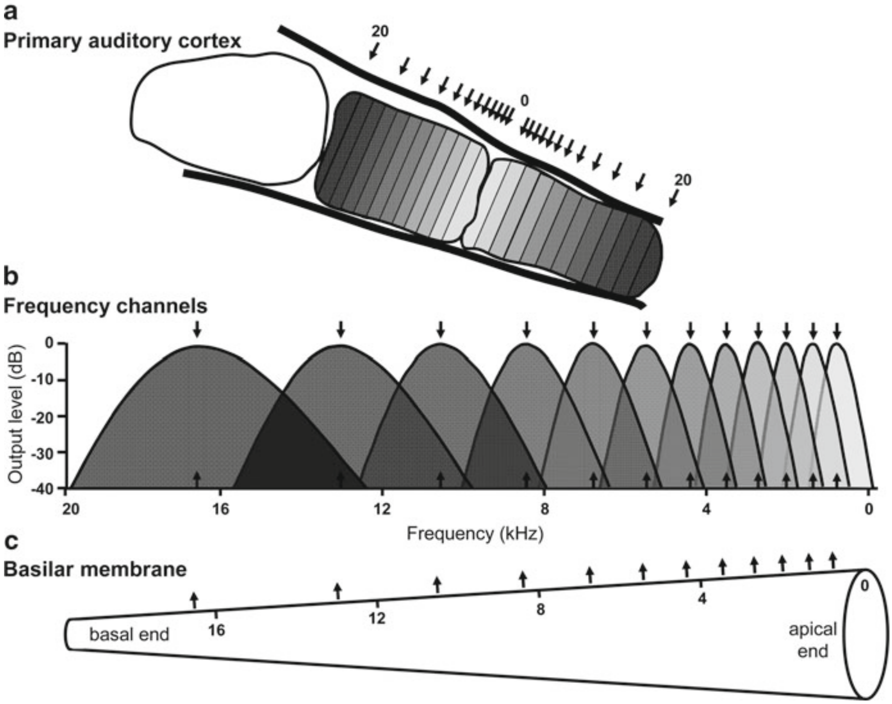


Fig. 7.2 (**a**) A highly schematic illustration of the basilar membrane in the cochlea as it might appear if it were unwound with the narrow, basal end being sensitive to high frequencies and the wide, apical end being sensitive to low frequencies. (**b**) A popular model of the cochlea in which the frequency selectivity of the basilar membrane is represented as an array of overlapping frequency channels. (**c**) A diagram showing the spatial organization of frequency coding in primary auditory cortex (fields Te1.0 and Te 1.1 on Heschl's gyrus). Within each field there is a systematic progression of isofrequency bands. The dark shading indicates high frequencies and the light shading represents low frequencies

auditory system and is found in all major auditory nuclei prior to auditory cortex. Within central auditory structures, the same gradient of frequency-specific coding is known as tonotopy. Numerous electrophysiological studies have recorded tonotopic responses in the mammalian auditory system. The best frequency of a neuron corresponds to the frequency at which the neuron is most responsive at low sound levels.

In primates, frequency selectivity has been shown to be greatest in primary auditory cortex with neurons becoming increasingly more broadly tuned in nonprimary regions of the belt and parabelt cortex (Morel et al., 1993). A prediction therefore is that the most convincing demonstration of human tonotopy should occur for primary auditory cortex rather than in nonprimary regions. Moreover, whereas single-frequency tones might be sufficient to stimulate primary auditory cortex, more complex sounds such as narrow-band noise bursts are preferable for investigating the response properties of surrounding areas.

7.2.1 Frequency Coding in Primary Auditory Cortex

At the advent of human neuroimaging, noninvasive measurements of electrical and magnetic field potentials were instrumental in documenting the tonotopic array in human auditory cortex (see Alain and Winkler, Chapter 4; Gabriel and Nagarajan, Chapter 5). The temporal acuity of these methods has been harnessed to accurately measure both transient (e.g., at sound onset and offset) and sustained (e.g., throughout the stimulus epoch) frequency-sensitive responses. From this early work, there is evidence that the latency of particular transient responses reflects the underlying tonotopy. The dominant peak in the evoked magnetic field that occurs around 100 ms after sound onset (the N100 or N1m) is a common marker for auditory cortical coding. For example, one study (Stufflebeam et al., 1998) demonstrated an increasing N1m latency as a function of decreasing stimulus frequency. This appeared to be a consistent finding in all five subjects tested for frequencies up to 1 kHz. The source of the frequency-sensitive activity has also been estimated using statistical methods to identify the location and orientation of the most likely dipole source. Dipole modeling of the transient evoked response has been applied to middle latency (10–50 ms) and longer latency (~100 ms) responses to single-frequency tones (e.g., Pantev et al., 1988, 1989, 1995), again with a high level of intra- and interindividual consistency. Within human auditory cortex, these results have suggested either a single tonotopic gradient (Pantev et al., 1988, 1989) or two mirror-image tonotopic gradients as depicted in Figure 7.2C (Pantev et al., 1995). In the case of a single frequency-sensitive gradient, the most commonly reported orientation is that of a high (medial) to low (lateral) axis, probably centered around Heschl's gyrus. Dipole modeling of the sustained response also supports the same interpretation (Pantev et al., 1996). More recently the focus of investigation has moved towards that of fMRI because it makes fewer assumptions about the underlying activity, rendering it more suitable for examining the spatial organization of fine-grained feature-specific coding in human auditory cortex (see Talavage, Johnsrude, and Gonzalez Castillo, Chapter 6). It is important to note that in fMRI the responses to an individual tone frequency cannot be measured directly. Instead, the response to a stimulus condition is compared to the response to a different stimulus condition. For example, to highlight regions most responsive to low frequencies, a low-frequency tone condition would typically be contrasted with a high-frequency tone condition. In terms of tonotopic mapping, it is important to clarify that this type of statistical contrast would not identify regions of low-frequency *specificity*, but would instead highlight regions with a preference for low-frequency sounds instead of high-frequency sounds. Nevertheless, this method is adequate for mapping out any loose tonotopic organization of the sort expected in human auditory cortex.

Some of the earliest fMRI studies to investigate tonotopy in human auditory cortex did not necessarily capitalize on the best spatial resolution achievable (e.g., Wessinger et al., 1997) and contrasted responses to only one low-frequency tone (55 Hz) and one high-frequency tone (880 Hz). More recent fMRI studies on tonotopy have addressed both of these issues. For example, Talavage et al. (2000) presented four pairs of narrow-band stimuli restricted to low (<660 Hz) and high (>2490 Hz) frequencies. Theirs was the first known fMRI study to have provided evidence for not one, but two frequency-dependent regions across Heschl's gyrus, and these shared a low-frequency border as in the primate core region. The locations of these frequency-dependent regions appear to correspond to areas Te 1.0 and Te 1.1 on the middle two-thirds of Heschl's gyrus (see Fig. 7.2C). All 12 hemispheres studied demonstrated these "mirror-image" tonotopic regions, with high frequencies being represented at the posteromedial and anterolateral endpoints and low frequencies at the common border in between.

Subsequently using a 3 Tesla scanner and four frequency-modulated tones each with different center frequencies (250 Hz–8 kHz), Schönwiesner et al. (2002) cast some doubt on the ability to convincingly demonstrate tonotopy using fMRI. Although the results obtained from this study showed very similar low- and high-frequency–dependent activation foci to those found by Talavage and his co-workers, the authors were uncertain about attributing them to two tonotopic maps because no systematic frequency–response gradients were observed and also because the foci lay on or near possible boundaries of other auditory fields.

Since the initial research by Talavage et al. (2000), at least three further human fMRI studies have identified two mirror-image tonotopic maps across Heschl's gyrus (Formisano et al., 2003; Talavage et al., 2004; Upadhyay et al., 2007). The study by Formisano and colleagues used an ultra-high-field (7 Tesla) scanner to measure responses to six tone frequencies (300 Hz–3 kHz). In the medial portion of Heschl's gyrus, their results documented a high (posteromedial) to low (anterolateral) frequency gradient that was reasonably consistent across the six listeners who participated in the study. The low-frequency response region shared a border with a second frequency gradient in the central portion of Heschl's gyrus, which further extended toward the anterolateral tip of the gyrus. In terms of the correspondence between these tonotopic maps and predictions about the underlying microanatomy, the medial gradient is consistent with the Te 1.1 and the central gradient is consistent with Te 1.0 (see Fig. 7.1C). Demonstrating tonotopy still remains a challenge, and not all recent fMRI studies have confirmed two mirror-image tonotopic maps (e.g., Langers et al., 2007a). This study found firm support only for a single gradient in Heschl's gyrus with a low-frequency response at the posteromedial end and a high-frequency response at the anterolateral end.

As a complementary approach to fMRI, the mapping of neuronal fiber projections provides another technique for examining the functional role of different auditory cortical regions. Diffusion tensor imaging (DTI) is a noninvasive MR method for identifying white matter fiber tracks and so is a useful way to investigate corticocortical connectivity. Upadhyay et al. (2007) used both imaging methods in a 3 Tesla scanner to reexamine tonotopy across Heschl's gyrus. The fMRI data confirmed the mirror-image fields on Heschl's gyrus. The DTI data revealed significant (isofrequency) projections between the two foci of high-frequency sensitivity and between the focus of low-frequency sensitivity and (non-isofrequency) projections between the high-frequency foci and their shared low-frequency border. Again, these projections are consistent with two core tonotopic fields.

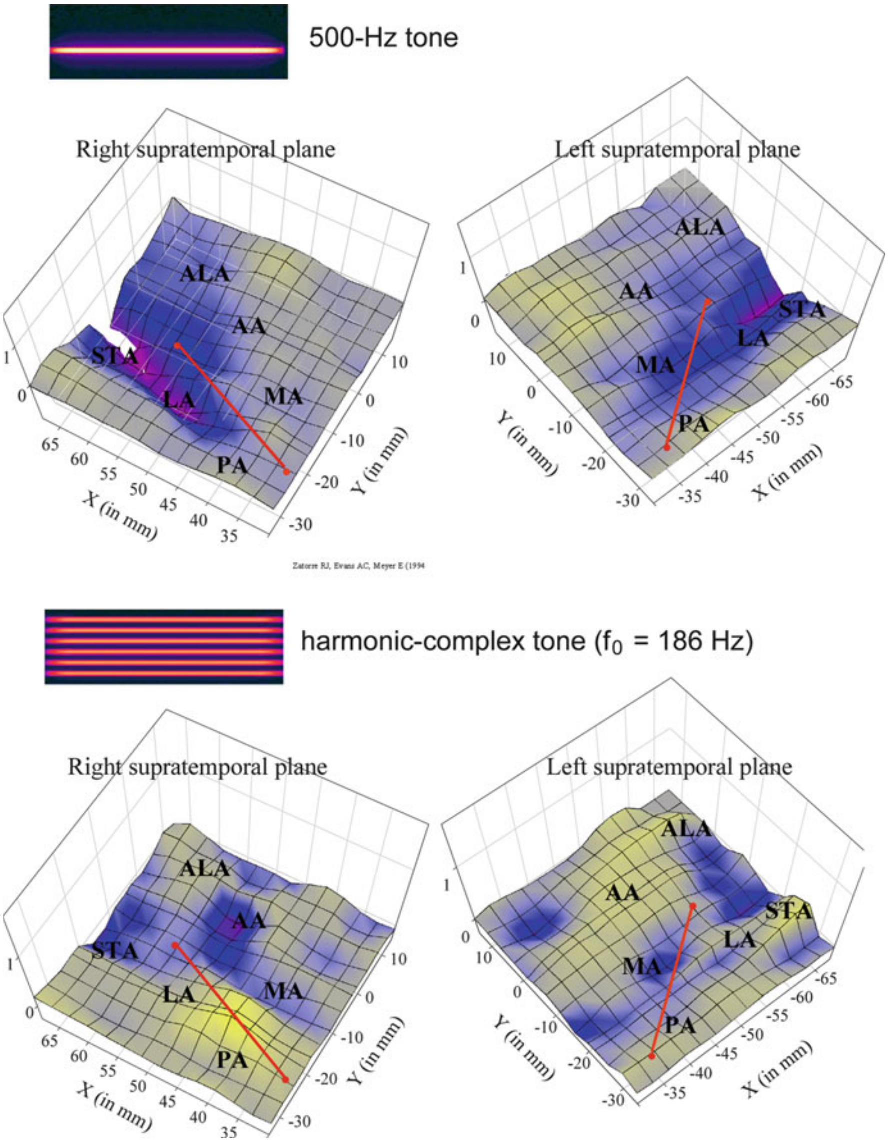
7.2.2 Frequency Coding in Nonprimary Auditory Cortex

In contrast with the general consensus of two mirror-image frequency gradients across Heschl's gyrus, the spatial arrangement of frequency sensitivity across nonprimary regions is less well defined. Talavage et al. (2000) postulated the existence of up to five nonprimary auditory fields, marked by four high-frequency and four low-frequency endpoints. Attributing these fields to cytoarchitectonic areas is somewhat dependent on the way in which the endpoints are "joined" up to form putative gradients and also on the parcellation scheme adopted. For example, in reference to the scheme shown in Figure 7.1B, one of these gradients could be located in the posterior area (PA), another in the anterior area (AA), and a third at the border of the supratemporal area (STA) and the lateral area (LA) (and so could be attributed to both or either field). Of course, without further evidence of a linear progression between the endpoints the interpretation of these data remains rather speculative and so the authors conducted a further study that used a technique of phase mapping to measure responses across a more complete range of frequencies (Talavage et al., 2004). Specifically, the stimulus in this experiment was a narrow bandwidth, amplitude-modulated noise with a center-frequency that was swept back and forth between 125 Hz and 8 kHz. The results confirmed tonotopy in four of the five nonprimary areas defined previously. The fifth region showed a broader-tuned response that was not sufficiently frequency selective to yield consistent results.

More recently, an fMRI study by Langers et al. (2007a) failed to provide reliable evidence of any tonotopically arranged fields outside primary auditory cortex, finding only small-scale variations in the optimal stimulus frequency in planum temporale. These authors concluded that frequency as an organizing principle was no longer obvious because at this stage in the auditory hierarchy, the sound signals were perhaps recoded to represent auditory scene analysis and auditory objects (see also Griffiths, Micheyl, and Overath, Chapter 8).

7.3 Broadband Signals

Another acoustic dimension associated with single frequency tones is that of signal bandwidth. Single frequencies form one endpoint of this dimension, while broadband noise forms the other. Bandwidth is therefore one of the most basic variables with which to characterize central auditory function. Broadband signals are generally more effective than single-frequency tones in evoking a neuronal response. This may be especially true in regions of nonprimary auditory cortex where single neurons respond more strongly to broadband stimuli than to single-frequency tones (Rauschecker et al., 1995). Several fMRI studies have demonstrated the large-scale consequences of this in terms of a relative increase in blood oxygen level–dependent (BOLD) activity across human auditory cortex for broadband signals. For example, Hall et al. (2002) compared activity for a single-frequency tone at 500 Hz and a harmonic-complex tone (f0 = 186 Hz, harmonics 1–5) that spanned 2.6 octaves.



500-Hz tone

Right supratemporal plane

Left supratemporal plane

Zatorre RJ, Evans AC, Meyer E (1994

harmonic-complex tone (f_0 = 186 Hz)

Right supratemporal plane

Left supratemporal plane

Fig. 7.3 A linear cut across the right and left supratemporal plane showing the spatial distribution of the response to the single-frequency tone (upper panel) and harmonic-complex tone (lower panel). The orientation of the long axis of Heschl's gyrus is plotted as a red line and the approximate central locations of the surrounding cytoarchitectonic fields are also shown. These data are reported in a different format in Hall et al. (2002)

They reported significantly more activity to the latter stimulus in Heschl's gyrus and in the lateral part of the supratemporal plane (Fig. 7.3). Comparing the peaks of activity with the architectonic scheme suggested that the increased activity by spectral cues might involve the fields LA and STA, as well as Te 1.2. These effects were significant at the group level and also showed good consistency across participants (i.e., for 5 of 6).

There are three possible functional interpretations for the observed growth in activity as a function of bandwidth. First, it is possible that the increase directly reflects the recruitment of neurons that perform spectral integration and thus have receptive fields that span large bandwidths. Conversely, it is also possible that the increase could be attributed to populations of neurons that each have a single best frequency and an excitatory response to sound, as this would lead to a spread of activity within tonotopic fields. These two explanations are rather difficult to separate using fMRI alone. The third explanation draws attention to sound level because it is an important acoustical feature that may contribute to the observed differences. Moreover, effects of both sound level and bandwidth have been found in overlapping regions of auditory cortex (Hall et al., 2001). Where details are reported, fMRI studies that manipulate bandwidth have sought to control for sound level by equating overall sound energy (e.g., Wessinger et al., 2001). It is likely that perceptual bases for matching, such as via a loudness model (e.g., Moore et al., 1997) would have a greater physiological validity at the cortical level, but this is unlikely to change markedly the current state of understanding about the effect of bandwidth on the pattern of auditory cortical activity.

7.4 Modulation

Natural sounds rarely contain acoustic features that are constant over time. Rather, they contain some kind of modulation over time either in frequency (FM) or in amplitude (AM). Important examples include animal vocalizations and species-specific communication signals. For humans, typically, slow-rate modulations (<50 Hz) are important for perceiving speech and recognizing melodies, while fast-rate modulations convey other types of sensations such as pitch and roughness. Common modulations in speech include frequency changes and formant transitions. These are complex sounds that contain multiple spectral peaks that sweep upwards or downwards in frequency over time, and also possess phonemic qualities. Further details about speech and music coding are provided by Giraud and Poeppel (Chapter 9) and Zatorre and Zarate (Chapter 10), respectively. To simplify their experiment, many investigators have chosen to present synthesized signals containing a single modulation component (e.g., sinusoidal amplitude modulation or a repeated train of noise bursts). It is those studies that are reviewed here.

In the auditory nerve, temporal modulation is represented faithfully in temporal discharge patterns (Joris & Yin, 1992). However, as one ascends the auditory system, neurons have an increasingly limited capacity to represent time-varying signals and so the temporal attributes of the signal become more indirectly represented by the neural code. This successive degradation in temporal precision is due partly to the temporal integration of inputs that occurs from one processing stage to the next and partly to the biophysical properties of neurons along the ascending pathway (e.g., Wang & Sachs, 1995). A good example of the cortical response to modulated signals is an electrophysiological study in marmoset monkeys (*Callithrix jacchus jacchus*; Lu et al., 2001). Results showed that cortical neurons in primary auditory cortex encode temporal modulation in terms of the temporal firing pattern and the mean firing rate, depending on the rate of modulation. Specifically, at slow modulation rates of up to 16 Hz, approximately 20%–55% of neurons coded the signal in an explicit manner, as a temporal discharge code. When the modulation rate exceeded 20 Hz, this proportion shifted to 20%–40% of neurons coding the signal in an implicit manner, using a discharge rate code. For the first time, this study highlighted the importance of the rate code for temporal information in the awake animal and it extended the range of the neural code to more closely match the wide perceptual sensitivities to low and high modulation rates. The rate code is highly relevant for fMRI because this method is more sensitive to changes in overall sustained discharge rate than to changes in neural synchrony (Logothetis, 2008).

7.4.1 Sustained and Transient Responses to Modulated Signals

fMRI studies have also shown that slow and fast modulation rates evoke different patterns of cortical activity, particularly in terms of its sustained and transient components. One of the early experiments to investigate this issue measured the response within a number of auditory structures to amplitude-modulated noise presented at rates of 4–256 Hz (Giraud et al., 2000). In auditory cortex, the preferred stimulus had a modulation rate of 4–8 Hz. This evoked the largest response and activity was sustained at a high level across the entire 30-s stimulus duration. In midbrain structures, such as inferior colliculus, a different pattern was observed. Here, the greatest response was to the noise modulated at 256 Hz and activity was restricted to the period immediately following stimulus onset (i.e., it was transient). The auditory cortical response to amplitude modulation has been more fully explored by Harms and Melcher (2002) and Harms et al. (2005). In these fMRI studies, stimuli were trains of noise bursts presented at rates of 1–35 Hz. There was a nonmonotonic relationship between rate and overall activity, with activity increasing from 1 to 2 Hz and then decreasing from 10 to 35 Hz. This can again be explained by the temporal envelope of the BOLD response over the 30-s stimulus duration. Activity was sustained for the slowest rates of modulation and then became more transient above 10 Hz (Fig. 7.4A). The authors suggested that the change to the shape of the BOLD response from sustained to transient with increasing modulation rate reflected the perceptual shift from individually resolved bursts (i.e., 1 and 2 Hz) to fused bursts (i.e., 10 and 35 Hz) forming a single “continuous” perceptual event. Activity was characterized separately for Heschl’s gyrus and the superior temporal gyrus, but appeared to be very comparable. The later study in 2005 demonstrated that the

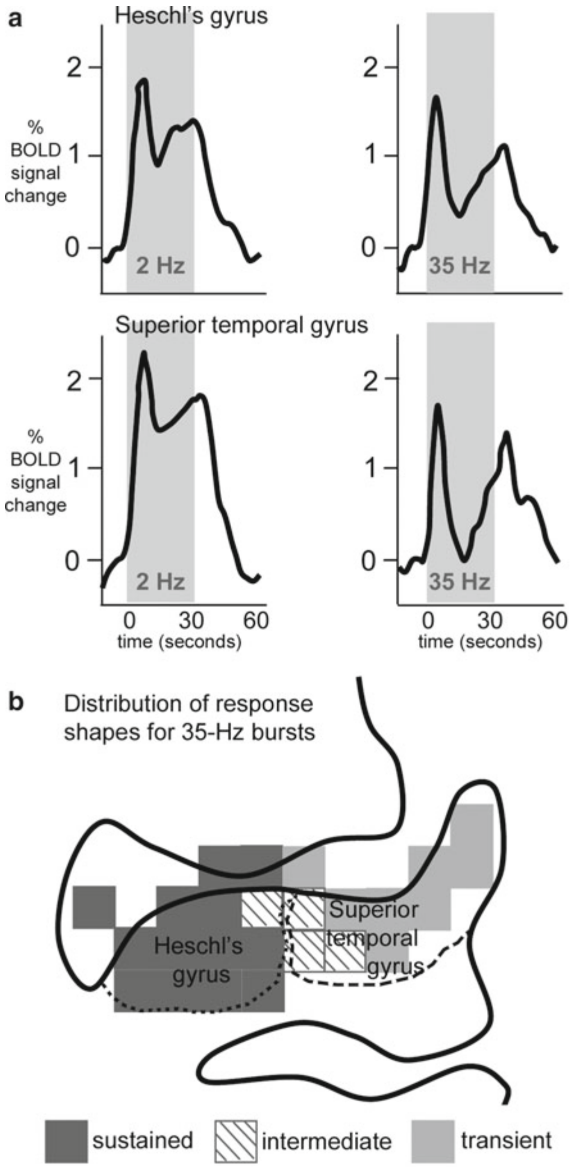


Fig. 7.4 (**a**) Temporal envelope of the fMRI response over the 30-s stimulus duration for slow (2 Hz) and fast (35 Hz) rates of modulation in Heschl's gyrus and superior temporal gyrus. (**b**) Single-subject example showing the distribution of response shapes for the 35-Hz burst rate in the left hemisphere. These schematic drawings are inspired by data reported in Harms et al. (2005)

transient response tended to be larger on the superior temporal gyrus than on Heschl's gyrus (Harms et al., 2005; see Fig. 7.4B), but the exact reason for this is unclear. It is possible that the larger amplitude of the transient response reflects the greater role of that region in segregating the auditory scene into distinct meaningful events (King & Nelken, 2009; Griffiths, Micheyl, and Overath, Chapter 8).

7.4.2 *Sensitivity to Slow-Rate Modulation Within Subdivisions of the Auditory Brain*

A number of fMRI studies have sought to identify which regions of human auditory cortex are most sensitive to slow-rate modulations. In the studies from the authors' laboratory, the signal was sinusoidally frequency modulated at a rate of 5 Hz and the stimulus for baseline comparison was a steady-state sound, matched in all other acoustic features. Hall et al. (2002) reported that the response to frequency-modulated tones occurred in Heschl's gyrus and in lateral parts of the supratemporal plane (possibly corresponding to regions LA and STA; Fig. 7.5). A particularly large response was seen just behind the lateral part of Heschl's gyrus in a region that might correspond to Te 1.2. The 2002 finding has since been replicated several times (e.g., Hart et al., 2003a, 2004). Dynamic ripples are synthesized sound stimuli that contain regular modulations in both amplitude and frequency. Responses to a range of such stimuli were measured by Langers et al. (2003) using fMRI. Compared to a noise baseline that contained no spectrotemporal modulation, these dynamic ripples were found to activate the posterior border of Heschl's gyrus and immediately behind on planum temporale. Consistent with the aforementioned results, Langers et al. (2003) found that most listeners who showed activity in lateral portions of auditory cortex did so best for slow modulation frequencies (i.e., 2 Hz), whereas voxels located more medially responded relatively well to faster rates of modulation (i.e., 32 Hz). This pattern is suggestive of a gross topography for modulation rate.

Of final note is another fMRI study that reported a disproportionately large response to upward and downward linear frequency sweeps in a large region posterior and lateral to Heschl's gyrus (termed T3; Brechmann et al., 2002). The previous modulation-related activity that was ascribed to Te 1.2 is broadly encompassed within area T3, although the borders of the different anatomical subdivisions differ. It is interesting to note that Brechmann et al. (2002) showed the modulation-related activity in this cortical region to be level independent. This finding suggests that the neural code for modulation in this nonprimary auditory cortical region perhaps reflects an abstract representation of the perceptual attribute of the stimulus. However, it has also been noted that this region appears to respond to other acoustic cues such as bandwidth (Hall et al., 2002), indicating no clear systematic segregation of response preference.

7.4.3 *A Common Representation of Modulation Rate?*

Although amplitude- and frequency-modulated sounds differ significantly in their spectral contents, they share the same modulation waveform that gives rise to their perceived time-varying properties. Until recently, it has been unclear whether cortical neurons might apply a common temporal processing mechanism to such a variety

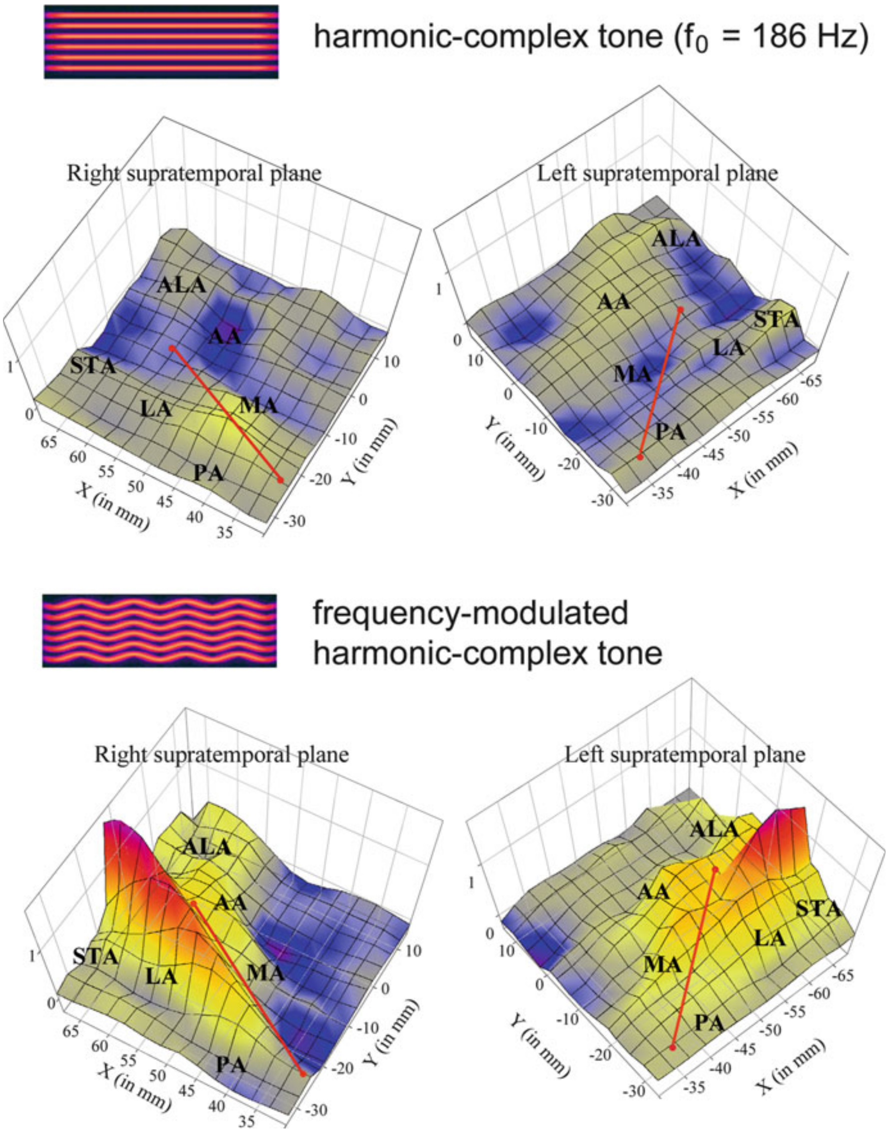


Fig. 7.5 A linear cut across the right and left supratemporal plane showing the spatial distribution of the response to the steady-state (upper panel) and frequency-modulated (lower panel) harmonic-complex tone conditions. The labels are the same as in Figure 7.3. These data are reported in a different format in Hall et al. (2002)

of time-varying signals. One way to answer this question is to measure systematically cortical responses to sinusoidally amplitude- and frequency-modulated signals, as these are two examples that are easy to manipulate and are representative of natural sounds. For instance, amplitude and frequency modulations are important components of communication sounds of animals and are found in a wide range of species-specific vocalizations including human speech. Psychophysical data have shown that listeners can make fine-grained judgments about phase differences when AM and FM signals are presented to separate ears (Saberi & Hafter, 1995). Although these results were consistent with the view that the auditory system might use a common neural code for both FM and AM information at relatively early neural stages (possibly before binaural convergence at the brain stem), at the time the physiological data were lacking. A number of studies since then have shed light on this matter. One relevant study reporting data recorded from single neurons in primary auditory cortex of awake marmosets was that by Liang et al. (2002). Electrophysiological recordings were made for both types of sinusoidally modulated stimuli presented at rates of 1–512 Hz. Results showed a high degree of similarity between cortical responses to both classes of stimuli. It was possible to identify a particular modulation frequency for which a neuron was selective, either by assessing its temporal firing pattern or its mean firing rate. Critically, this selectivity was shown to be similar regardless of whether the temporal modulation was created in the amplitude or frequency domain.

A comparable study in human auditory cortex has been conducted using fMRI to measure sustained cortical responses to signals that were modulated at a rate of 5 Hz in the time domain and separately in the frequency domain (Hart et al., 2003a). In this study, two carrier signals were used to provide some internal validation of the effects; a single-frequency tone and a harmonic-complex tone, both with F0 = 300 Hz. When compared with their matched steady-state carriers, both types of modulation evoked significantly greater activity in the lateral portion of Heschl's gyrus (possibly Te 1.2) and in adjacent parts of the planum temporale (possibly LA and STA), replicating the previous findings. The most important finding was that the two activation patterns were largely overlapping, supporting the view of a common neural code. In summary, these results indicate that cortical neurons extract the temporal profiles of modulated tones by the same mechanism, regardless of the spectral content of the sounds. Results from this human fMRI study suggest that this function is not restricted to primary auditory cortex (i.e., Te 1.0 and 1.1).

Generally speaking, the auditory steady-state response (aSSR) is the main approach to magnetoencephalography (MEG) and electroencephalography (EEG) measures of AM and FM coding in human auditory cortex. The aSSR is an elicited response that has the same frequency as the corresponding stimulus modulation frequency. Luo et al. (2007) investigated coding transitions as a function of stimulus rate dynamics using MEG. Their studies manipulated the modulation rates for AM and FM using signals that contained simultaneous AM (fixed at a modulation rate of 37 Hz) and FM (varying in modulation rate from 0.3 to 30 Hz). Results were again consistent with the view of a population of neurons (or a paired set of populations) in auditory cortex that co-encode independent AM and FM stimulus modulations in a naturally grouped manner.

7.5 Sound Level

A range of scales are available for measuring sound level. A common objective measure of sound level ("intensity") is the decibel (dB) scale, which relates to the power of the sound energy. Decibels represent the ratio of a given intensity (10^x watts/m^2) to the standard threshold of hearing, so that the threshold of hearing corresponds to 0 dB. However, listeners do not describe sounds in terms of decibels, but instead use language such as "soft" or "loud." Intensity and loudness are measures of different sound level characteristics. Two different 60-dB sounds will rarely have the same loudness because the judgment of loudness takes into consideration the ear's sensitivity to the component frequencies of the sound. A common "loudness" scale is that measured in phons. The basis for the phon scale references each sound to the equivalent decibels level for a 1-kHz tone. Thus if a given sound is judged to be as loud as a 1-kHz tone at 60 dB, then it is said to have a loudness of 60 phons. For broadband signals, the loudness is determined by the auditory excitation pattern, integrated across frequency (Moore et al., 1997).

Like frequency, level is one of the most basic attributes of sound and is coded at the first stage of cochlear transduction. At the auditory periphery, sound level is represented by the firing rates of neurons at the center of the excitation pattern (e.g., Liberman, 1978), by the spread of the excitation pattern (e.g., Chatterjee & Zwislocki, 1998), and by temporal synchrony in the pattern of neural firing (e.g., Brosch & Schreiner, 1999). The dynamic range of human hearing is extremely broad and yet is exquisitely sensitive to discriminating very small changes in pressure variations in the air across this range (Viemeister & Bacon, 1988). At 1 kHz, the lowest detectable sound pressure level is about 10^{-12} watts/m^2. This corresponds to 0 dB SPL (decibels sound pressure level). Arguably, the highest sound level that can be tolerated without causing intense pain and cochlear damage is about 10^{13} watts/m^2 (120 dB SPL). Although the dynamic range of hearing exceeds 100 dB, individual auditory neurons are sensitive to a much narrower range of levels (generally 20–30 dB). Sensitivity to sound level is improved because different neurons adjust their input–output functions according to the prevailing distribution of levels (Dean et al., 2008).

Neurophysiological studies in animals indicate that sound level may be represented by neurons that are distributed within populations that subserve other functions, including the sharpness of frequency tuning to pure tones (Recanzone et al., 1999). At low sound levels, activated neurons show sharp frequency tuning close to the stimulating frequency, but at higher intensities of the same tone frequency there is a spread of excitation to neurons with characteristic frequencies both higher and lower than the stimulating frequency (Phillips et al., 1994). Within auditory cortex, the response of the neural population to sound level becomes highly complex. Temporal coding has largely disappeared and rate coding is a mixture of both monotonic and nonmonotonic neuronal responses to increasing sound level. Monotonic units typically show a progressive increase in discharge rate as a function of sound level, although a maximum firing rate can be reached above which further increases in sound level have no effect. In contrast, nonmonotonic units are those for which further increases in sound level result in a progressive decrease in activity from the maximum value. In other words, nonmonotonic units are tuned to particular best SPLs. Monotonic rate-level functions appear to be in the substantial majority throughout the central auditory system, at least for broadband noise stimuli (Phillips et al., 1985). Thus, perhaps one might predict that the neuroimaging response to broadband noise should also show monotonic dependencies on sound level, as these techniques provide an indication of the summed activity of a neural population. For single-frequency tones, the predictions become less clear because there is a high proportion of nonmonotonic rate-level functions in auditory cortex (Phillips et al., 1985, 1994). For single-frequency tones, neurons showing monotonic and nonmonotonic behavior will contribute substantially to the level dependence of cortical activity. Human neuroimaging studies have therefore taken an exploratory approach to characterizing the predominant relationship between sound level and amount of sound-related activity using different stimuli and different measures of sound-related activity.

7.5.1 Monotonic Level-Dependent Functions in Human Auditory Cortex

EEG/MEG studies have reported an effect of increasing sound level on various parameters of the human auditory evoked response including an increase in the N100/N100m amplitude, a reduction in the N100/N100m latency and an increase in the N1–P2 peak-to-peak amplitude (Stufflebeam et al., 1998). fMRI and PET have also been used to measure sound level–related activity and results have similarly indicated a growth in activity with increasing sound level across human auditory cortex. Not all studies have the sensitivity to determine the shape of the level-dependent function. Some have been somewhat limited by their narrow sampling of the full dynamic range and their choice of large step sizes (e.g., Jäncke et al., 1998; Lasota et al., 2003). In studies that have used a more optimal parametric design, the extent of activation and response magnitude both tend to increase monotonically. One exception is the PET study reported by Lockwood et al. (1999) in which regional cerebral blood flow (rCBF) for a 500-Hz tone showed a somewhat U-shaped function. As a more representative example, Figure 7.6 illustrates data reported by Hart et al. (2002) for a 300-Hz tone. Analysis confirmed that the number of activated voxels in auditory cortex was significantly determined by sound level across the 42- to 96-dB SPL range. Such a pattern was observed in both hemispheres, but was strongest in the hemisphere contralateral to the monaural stimulus. Moreover, on this contralateral side, the growth was particularly sharp at the highest sound levels. Typically, the level-dependent function continues its upward trajectory even at intense sound levels. The response seems to show no evidence of nonmonotonicity or of reaching a plateau. Similar results have been reported for a range of different sound stimuli, including a 300-Hz tone that presented up to 96 dB SPL

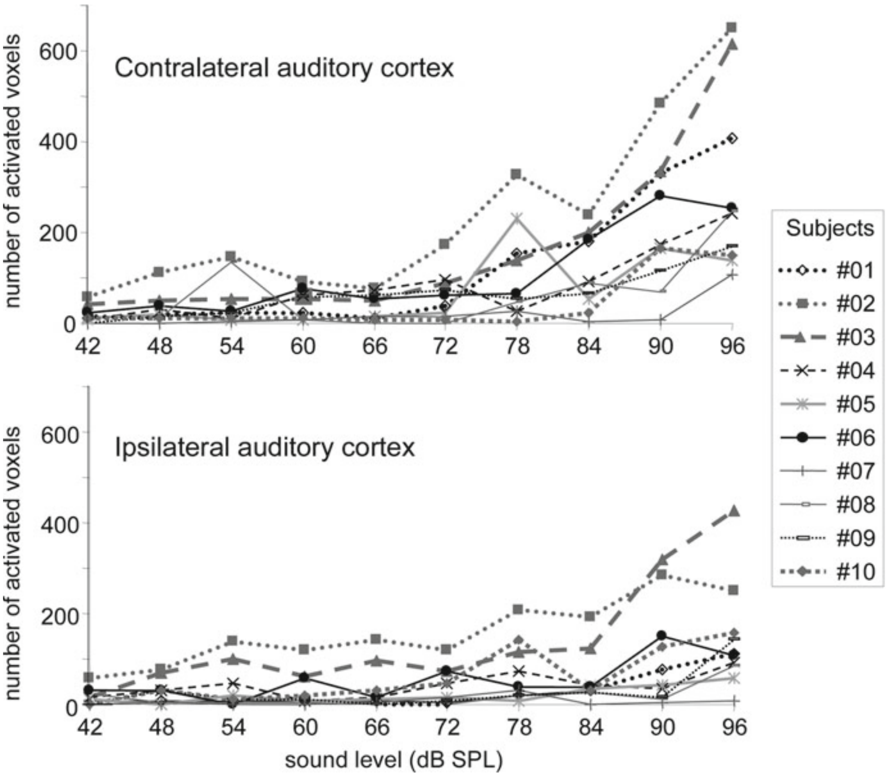


Fig. 7.6 An example of the systematic changes in auditory cortical activity as a function of sound level, in response to a 300-Hz tone. To be classed as "activated," voxels had to reach a significance threshold of $p < 0.001$. The number of activated voxels was calculated separately for each sound level contrast (i.e., tone – silent condition) for each of 10 normal-hearing subjects. (A version of this figure was presented at the 24th Association for Research in Otolaryngology MidWinter Meeting, 2001, St. Petersburg, Florida, USA. The group means are published in Hart et al., 2002.)

(Hart et al., 2002); two frequency-modulated tones spanning the spectral range 0.5–1.0 kHz and 4–8 kHz presented up to 80 dB sensation level (Langers et al., 2007b); a 4.75-kHz tone presented up to 96 dB SPL (Hart et al., 2003b); a 4-kHz tone presented up to 90 dB SPL (Lockwood et al., 1999); and a continuous broadband noise presented up to 99 dB SPL (Sigalovsky & Melcher, 2006). The rate of growth as a function of sound level does not appear to be the same across all frequencies. In a study that directly compared the effect of two tone frequencies, Hart et al. (2003b) demonstrated that, within Heschl's gyrus, the response to a low-frequency tone was flat between 42 and 66 dB SPL and then showed a rapid growth that continued up to the highest level studied (96 dB SPL). In contrast, the response to a high-frequency tone increased steadily across the same range of levels. These results concur with physiological evidence suggesting that recruitment of primary auditory cortical neurons may be different at high and low frequencies (Phillips et al., 1994).

Systematic increases in both extent and magnitude of the response do not always co-occur in the same data set. For example, for syllables and pure tones presented at levels of 75, 85, and 95 dB SPL, Jäncke et al. (1998) found a significant increase in the extent of auditory cortical activity, but no significant effect on response magnitude. However, for monosyllabic words presented at levels from 65 to 110 dB, Mohr et al. (1999) found a reliable increase in response magnitude, but not extent. Comparable outcomes for extent and magnitude might be expected because, at a simplistic level of interpretation, growth with sound level is physiologically consistent with a regional increase in the general activity of the underlying neuronal population. A dissociation between the shape of the level-dependent function for extent and magnitude might simply reflect lack of sensitivity in the (BOLD or rCBF) neuroimaging measure. Indeed, it has been suggested that extent is perhaps a less reliable measure of activation than magnitude (Mohr et al., 1999; Hall et al., 2001), especially in experiments with many stimulus conditions. An alternative explanation, especially in those studies utilizing fine spatial resolution, is that a dissociation between the extent and magnitude measures might represent either neural recruitment or a local increase in neural activity, respectively. The preceding discussion has hopefully emphasized the point that comparisons between animal and human data on level sensitivity are unlikely to be straightforward. Although it is reasonable to anticipate neural recruitment for high sound levels (see Hart et al., 2002), increases in BOLD/rCBF responses are not necessarily indicative of increases in neural firing rate, especially given the contribution of nonmonotonic units to sound level coding. At the cortical level, there are profuse local inhibitory influences (Manunta & Edeline, 1998; Logothetis, 2008), although a direct local contribution to the observed nonmonotonicity of rate-level functions has yet to be demonstrated. Nevertheless, if nonmonotonic responses are mediated by summation of excitatory and inhibitory inputs to cortical neurons, an *increase* in subthreshold activity at high sound levels would occur despite the *reduction* in the output from such units. The greater metabolic demand caused by such a rise in synaptic activity would most likely be responsible for an *increase* in the BOLD/rCBF response (Logothetis, 2008).

7.5.2 Sensitivity to Sound Level Within Subdivisions of the Auditory Brain

At every major stage of the ascending auditory pathway, significant rate-level functions have been demonstrated in humans. To our knowledge, only one fMRI study has so far quantified level-dependence of activation within subcortical auditory structures (Sigalovsky & Melcher, 2006). Using a broadband continuous noise stimulus presented binaurally at 30, 50, and 70 dB sensation levels (equivalent to 50–99 dB SPL), the main trend was again one of a monotonic increase in activity. This pattern was observed in the cochlear nucleus, superior olivary complex, inferior colliculus, and medial geniculate body (and auditory cortex).

A small number of neuroimaging studies have distinguished level-dependent functions in different anatomically and functionally distinct subdivisions of human auditory cortex. One of the first fMRI studies to investigate this issue was conducted by Hart et al. (2002). These authors quantified the response to sound level within three anatomically defined regions of human auditory cortex: (1) Heschl's gyrus (the medial two-thirds probably incorporating the primary fields Te 1.0 and Te 1.1), (2) the anterior lateral area (representing Te 1.2), and (3) planum temporale (possibly including LA, STA, and PA). Within these three regions, Hart and colleagues plotted the proportion of suprathreshold ($p < 0.001$) voxels and the mean scaled percent signal change as a function of sound level. In this study, the range of sound levels spanned 42-96 dB SPL in 6-dB steps and the stimulus was a 300-Hz tone. Of the three anatomically defined regions, the response centered on Heschl's gyrus was the most sensitive to increasing sound level for both magnitude and extent measures of activity. Consistent with this finding was a subsequent fMRI study demonstrating a monotonic increase in the percentage of voxels within Heschl's gyrus that reached the chosen threshold of $p < 0.0001$ (Lasota et al., 2003). This study used a 1-kHz tone presented at a range of sound levels (0–50 dB hearing level). Langers et al. (2007b) also commented that Heschl's gyrus was the dominant source for their sound-level dependencies.

Although not specifically commenting on putative differences between cortical regions in their sensitivity to level, Sigalovsky and Melcher (2006) examined four regions of interest that defined broad subdivisions of auditory cortex. (1) The posteromedial two-thirds of Heschl's gyrus was intended to approximate Te1.0 and Te 1.1; (2) the remaining antero-lateral third of Heschl's gyrus was probably equivalent to Te 1.2 (as shown in Fig. 7.1); (3) the entire planum temporale was assumed to incorporate lateral belt regions (LA, PA, and STA); and (4) an anteromedial region, located in front of Heschl's gyrus up to the circular sulcus, was possibly the human homologue of medial belt regions (MA and AA). The authors applied a number of independent measures of sound-related activity. The primary "magnitude" analyses first identified voxels reaching significance (at $p < 0.01$) and then across subjects and hemispheres calculated the average maximum percent change at the onset of the noise stimulus (relative to a silent baseline) and the average maximum percent change at the offset of the noise across each sound level condition. A supplementary "extent" analysis counted numbers of voxels within the region of interest that exceeded a probability of activation ($p < 0.01$). Comparing the 30- and 70-dB conditions, there was an increase in both the onset and offset percent change in all of the subdivisions except the anterior medial nonprimary auditory cortex where the same trend did not reach significance. However, this region was generally less responsive to sound stimulation than the other cortical regions. Again, the most significant level-dependent change occurred in primary auditory cortex, albeit for the magnitude of the offset response, not the onset response.

7.5.3 Searching for a Topographic Representation of Sound Level

In the mammalian primary auditory cortex, an orderly spatial organization of a number of parameters related to the encoding of sound level has been demonstrated. Organizing principles include minimum threshold, dynamic range, best SPL, and nonmonotonicity of intensity functions. The analysis of several neuroimaging data sets has explored the evidence for a systematic relationship between sound level and the location of auditory activity (ampliotopy). On balance the results are somewhat negative because they fail to demonstrate ampliotopy (see Hart et al., 2002; Sigalovsky & Melcher, 2006). This does not necessarily rule out the possibility that ampliotopy does exist. It may simply remain obscured by current measurement techniques.

7.5.4 A Physical or Perceptual Representation of Sound Level?

Given the range of scales available for measuring sound level, Hall et al. (2001) considered the issue of control for sound level in the context of comparing auditory cortical activity for single-frequency tones and broadband signals. If intensity is fixed while signal bandwidth is increased, then loudness nevertheless increases because the signal spans a greater number of frequency channels. The question therefore arises, “should one match stimuli for intensity or loudness?” To address this, Hall et al. (2001) presented a range of single-frequency tones and harmonic-complex tones that were matched either in decibels or phons. When the fMRI data were collapsed across stimulus class, neither activation extent nor magnitude significantly correlated with the decibel scale. In contrast, both extent and magnitude correlated significantly with the phons scale. On the basis of these results, the authors speculated that loudness may be an important aspect of the auditory cortical representation of sound.

More recently, Langers et al. (2007b) considered auditory cortical responses as a function of intensity and loudness using low- and high-frequency stimuli presented across a 70-dB range, in steps of 10 dB. To address whether intensity or loudness was the main characteristic driving the pattern of level-dependent activation, the authors compared two groups of listeners, one with normal hearing and one with age-related sensorineural hearing loss. This type of impairment reduces high-frequency hearing sensitivity and is accompanied by loudness recruitment at high frequencies (a disproportionate rise in loudness ratings as a function of intensity). If loudness were the driving factor, then a dissociation would be predicted between decibels and equivalent loudness curves across the two groups of participants at high frequencies. Generally, the fMRI results revealed monotonic increases in the magnitude of activation across intensity and loudness. At low frequencies, the steepness of the intensity- and loudness-dependent functions did not differ across the hearing impaired and normal hearing groups. This was also true at high frequencies for the loudness-dependent function. However, at high frequencies the intensity-dependent function was significantly steeper in the hearing impaired group than in the group with normal hearing (mean slope was 37 and 21 × 10^{-3}%/dB, respectively). These results therefore support the conclusion that loudness relates more strongly to cortical activation than does intensity. This interpretation is also consistent with the general view that cortical activation reflects the correlate of the subjective strength of the stimulus percept.

7.6 Pitch

Pitch is one of the most fundamental auditory percepts. It can be defined in musical terms by any sound that can be used to produce a melody, and can be ordered on a scale from low to high. Pitch plays an important role in music perception and in language (conveying prosody and, in some languages, semantic information). Pitch is a perceptual attribute of sound, but it is determined by physical characteristics of the acoustic signal including its frequency (e.g., in the case of single-frequency tones) or its temporal periodicity (e.g., in the case of complex sounds). These two physical cues form the basis of two mechanisms for the neural coding of pitch: a rate-place code and a time code (de Cheveigné, 2005). Harmonic-complex tones are an interesting example because depending on whether their frequency components are "resolved" or "unresolved," the pitch can be conveyed by either, or both, neural codes. Defining each harmonic as "resolved" or "unresolved" depends on its neural activation pattern within the peripheral auditory system. The low-numbered (resolved) harmonic components tend to fall within individual frequency channels, producing a characteristic excitation pattern across the membrane in which there is a one-to-one mapping between the spectral peaks in the acoustic signal and the peaks of excitation. The sensation of pitch could therefore arise from a detection of the harmonically related, resolved peaks of neural activity. This is the rate-place code. Although it is still debated at what point the harmonics cease to be resolved along the basilar membrane, it is generally accepted that in a harmonic series those components below the seventh are resolved and those above the 13th are unresolved (Houtsma & Smurzynski, 1990). The unresolved harmonics are not individually represented on the membrane, but instead multiple harmonics fall within a single frequency channel and the resulting excitation pattern contains no distinct spectral peaks. The pitch of these stimuli can be determined instead from the output of a single channel containing many interacting harmonics, whose repetition rate corresponds to the F0 (i.e., the pitch) of the complex tone (Houtsma & Smurzynski,1990; Carlyon et al., 1992). This is the time code.

Although pitch processing mechanisms most probably exploit both spectral and temporal information (e.g., Carlyon et al., 1992; de Cheveigné, 2005), many neuroimaging investigations have sought to eliminate the spectral cues for pitch to

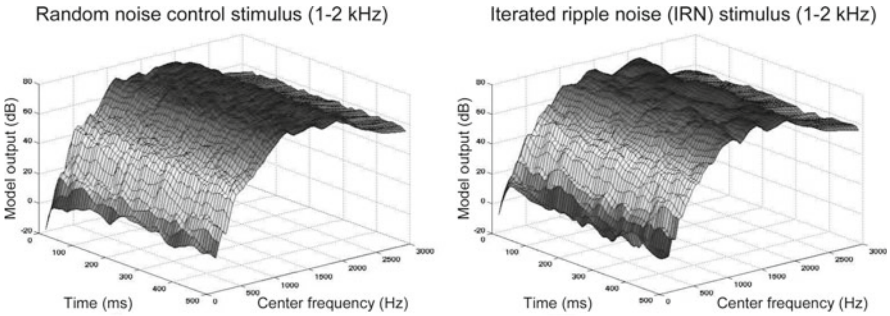


Fig. 7.7 Simulated output of the cochlea in response to a random noise stimulus and to an iterated ripple noise (IRN) stimulus. The model output in decibels is plotted as a function of time and of the center frequency of each auditory frequency channel (or each place in the cochlea) across a bandwidth of 1–2 kHz. Note that the spectral content is comparable across the two signals since the cues for pitch are conveyed in the temporal dimension of the IRN stimulus. (Courtesy of C. J. Plack.)

isolate the neural representation of the time code. Stimuli for which the dominant cue for pitch is temporal rather than spectral include unresolved harmonic-complex tones, amplitude-modulated tones, regular interval sounds and dichotic pitches (Fig. 7.7). For these stimuli, pitch cues are not carried in the spectral (i.e., tonotopic) pattern of neural activity and pitch coding may therefore engage additional regions of the auditory cortex that are not so sharply tuned to frequency. One popular type of regular interval sound is iterated ripple noise (IRN). IRN is created by generating a sample of random noise, delaying it, and adding or subtracting the duplicate to or from the original (Yost, 1996). The pitch of an IRN is equivalent to the reciprocal of the delay imposed. The pitch strength (salience) can be increased by increasing the number of delay-and-add iterations (Yost et al., 1996). Both pitch value and strength can be manipulated in a systematic manner, with little effect on the spectral content of the stimulus.

7.6.1 *Pitch Sensitivity within Subdivisions of the Auditory Brain*

One way to identify pitch-sensitive activity is to compare the response to IRN with that to a random noise signal that has the same spectral content. When Patterson et al. (2002) contrasted a sequence of IRN bursts with a fixed pitch and a sequence of random noise bursts, they found activation in lateral Heschl's gyrus. This result was consistent in eight of the nine listeners. The putative anatomical field corresponding to this region is Te 1.2 (see Fig. 7.1C). A number of other PET and fMRI studies provide convergent evidence that lateral Heschl's gyrus is maximally responsive to IRN. Moreover, two of these studies have demonstrated a systematic increase in the response within lateral Heschl's gyrus as a function of increasing pitch strength (Griffiths et al., 1998; Hall et al., 2005), as shown in Figure 7.8. This relationship was examined using IRN signals in which the number of delay-and-add iterations ranged from 0 to 16.

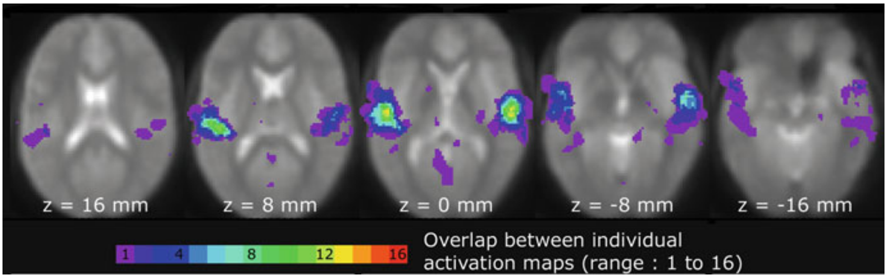


Fig. 7.8 An incidence map showing auditory cortical increases in activity as a function of pitch salience (an increase in activity for IRN with 0, 1, and 16 add-and-delay iterations). The color code illustrates the variability of the effect across 16 listeners. All maps are overlaid onto the same 5 horizontal brain images (z = +16 to –16 mm) in neurological convention (i.e., left = left). (The original version of this figure is published in Hall, Barrett, Akeroyd, & Summerfield. [2005]. *Journal of Neurophysiology*. Cortical representations of temporal structure in sound. 94(5), 3181–3191. doi: 10.1152/jn.00271.2005. http://jn.physiology.org/content/94/5/3181.full.pdf+html. This is an unofficial adaptation or translation of an article that appeared in a publication of the American Physiological Society. The American Physiological Society has not endorsed the content of this adaptation or translation, or the context of its use.)

If this region is to be called a "pitch center" then it should represent subjective pitch regardless of the spectral, temporal, or binaural characteristics of the stimulus (e.g., Bendor & Wang, 2005). One fMRI study filtered harmonic-complex tones into low and high spectral regions to produce resolved complex tones evoking a strong sense of pitch and an unresolved complex tone evoking a weak sense of pitch (Penagos et al., 2004). Contrasting these two stimulus conditions again revealed patches of activity around lateral Heschl's gyrus. The amplitude of the BOLD response was significantly smaller for the weak pitch condition than the strong pitch condition. In a recent fMRI study, Hall and Plack (2009) measured cortical responses to seven different pitch-evoking stimuli, each with different spectral and temporal characteristics (pure tone, resolved and unresolved harmonic-complex tones, a wideband harmonic-complex tone, a binaural pitch stimulus [Huggins pitch], and two types of IRN). The results for the IRN stimulus showed good agreement with previous studies. However, a different pattern of activation was reported for the other five pitch-evoking stimuli. Instead of lateral Heschl's gyrus, planum temporale was most consistently activated across listeners. However, even in this region there was a high degree of individual variability (illustrated in Fig. 7.9). From this subset of six listeners, three showed planum temporale activity for many of the pitch stimuli presented but for three other listeners activity was located elsewhere. This finding would indicate that it is rather premature to assign special status to lateral Heschl's gyrus solely on the basis of activation patterns. A recent fMRI study used a novel form of group analysis to explore the cortical representations of pitch and sound objects (Staeren et al., 2009). Stimuli were chosen from four different sound categories (complex tones, singers, cats, and guitars) and each contained examples at three different pitch values (250, 500, and 1000 Hz). Responses that

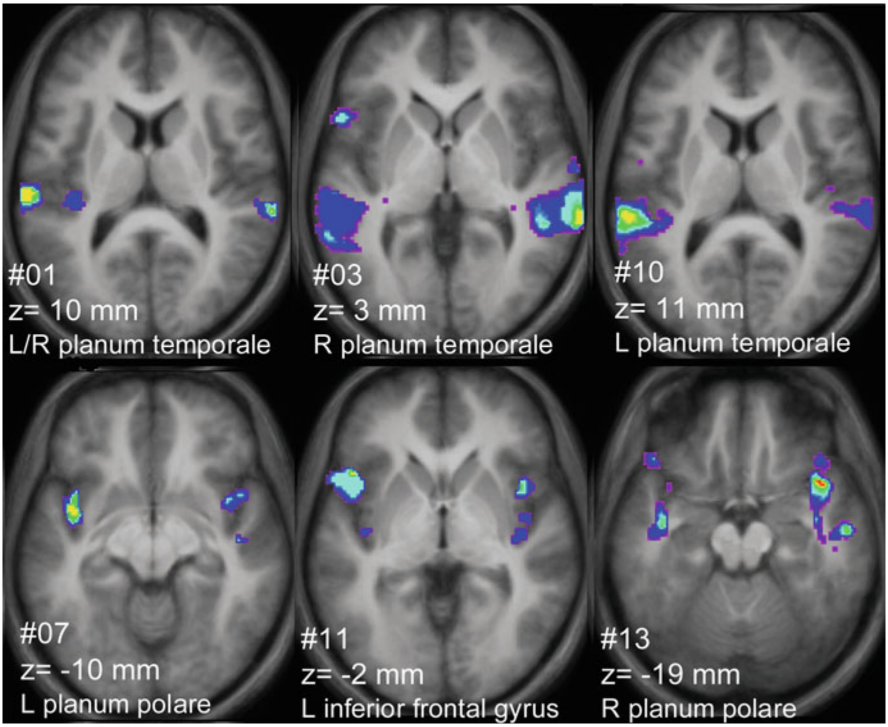


Fig. 7.9 Incidence maps showing the consistency of pitch-related activation for five pitch stimuli presented to six different listeners. Activity was calculated separately for each pitch contrast (i.e., pitch – noise condition) using a significance threshold of $p < 0.01$. For each listener, the activity maps were combined and the resulting color coding indicates how many of the pitch stimuli evoked activity at a particular voxel (blue = 1, cyan = 2, green = 3, yellow = 4, red = 5). All maps are overlaid onto the individual anatomical brain image in neurological convention (i.e., left = left). Group mean data are reported in Hall and Plack (2009)

discriminated between the pitch values were distributed across patches of posterolateral Heschl's gyrus and planum temporale, in accordance with previous measures of pitch-related activity. At the time of writing, the search for a generalized human pitch center is ongoing.

7.6.2 Pitch Onset

Neuroimaging investigations of pitch processing have typically presented sequences of bursts of pitch-evoking stimuli separated by intervals of silence. Neural responses to the control condition (e.g., a sequence of random noise bursts) are subtracted from the pitch condition, with the residual activation identified as the 'pitch-specific' response. It is well known that many auditory cortical neurons are highly responsive at stimulus onset (e.g., Lu et al., 2001; Liang et al., 2002) and so one might therefore expect a large transient energy response at each sound onset for these stimulus sequences. It is possible that neuroimaging measures have confounded pitch onset and energy onset responses. However, careful design of the stimulation paradigm is able to separate out the transient response to the pitch onset from that to energy onset (e.g., Krumbholz et al., 2003; Chait et al., 2006). In the continuous stimulation paradigm, bursts of pitch-evoking stimuli are introduced into an ongoing noise signal, thus removing the changes in energy at the transition from baseline to pitch. Further, the temporal resolution of EEG and MEG is ideally suited to isolating the transient onset responses. Using this paradigm in the context of an MEG study, Krumbholz et al. (2003) found a positive deflection with a latency of about 150 ms at the transition from random noise to IRN. Such a deflection was not seen for the transition from IRN to random noise and so it was termed the ‘pitch onset response.’ In addition, the amplitude of the pitch onset response increased with increasing pitch strength and the latency of the pitch onset response decreased as F0 increased. Crucially, the pitch onset response appears to be consistent across different types of pitch-evoking stimuli because a similar pattern of results has been obtained for both a tone-in-noise and a binaural (Huggins) pitch (Chait et al., 2006). The continuous stimulation paradigm is not limited to EEG and MEG, but has also been implemented in fMRI. Using a harmonic-complex tone and a complex Huggins pitch, Garcia et al. (2010) demonstrated enhanced sensitivity to pitch compared to a more ‘classic’ stimulation paradigm that alternated bursts of pitch-evoking stimuli with short periods of silence. Results indicated that Heschl’s gyrus was most engaged by the changes in sound energy, whereas pitch information was best represented in parts of planum temporale.

The neural generators of the pitch onset response have also been estimated using dipole source modeling using EEG and MEG data (e.g., Krumbholz et al., 2003; Chait et al., 2006). According to these results, the source is typically located close to Heschl’s gyrus but, given the rather poor spatial resolution of these methods, other methods may be more informative. Depth-electrode recordings in patients who are candidates for epilepsy surgery *do* allow for more accurate localization of the stimulus-evoked electrical signals because those measures allow for direct localization without source modeling. Two recent studies have presented IRN in the context of the continuous stimulation paradigm to patients undergoing surgery (Schönwiesner & Zatorre, 2008; Griffiths et al., 2010). In the single case reported by Schönwiesner and Zatorre (2008), a depth electrode was directed within the lower bank of the Sylvian fissure about 5 mm behind Heschl’s gyrus running parallel to it, so that five of the nine electrode contacts recorded electrical activity from this gyrus. Contacts 2 and 3 (close to the medial two-thirds of Heschl’s gyrus) responded strongly to the energy onset response, while contact 5 (on the supratemporal plane close to lateral Heschl’s gyrus) responded best to the pitch onset. In the study by Griffiths et al. (2010), a depth electrode was implanted along the long axis of Heschl’s gyrus in one hemisphere. In both patients, significant pitch-related responses were recorded between contacts 2 in medial Heschl’s gyrus (Te 1.1) and 10 in central Heschl’s gyrus (Te 1.0). After the transition from noise to IRN, there was a sustained increase in power for the oscillatory activity in the high gamma range (80–120 Hz). Griffiths et al. suggested that this induced gamma response is related to the perception of pitch because it was found to be specific to those IRN stimuli evoking a sensation of pitch (i.e., 128 and 256 Hz) and not for those IRN stimuli that did not evoke pitch (i.e., 8 and 16 Hz). In general conclusion, a continuous stimulation paradigm would appear to improve specificity of pitch-related activity by eliminating activation related to energy onset.

7.6.3 Listening to Melodies

When different pitches are presented in a temporal sequence, they form a melody. Melody plays a critical role in music perception and in the recognition of familiar tunes. In terms of the stages of sound processing, melody perception can be construed as one of the highest levels. Functional neuroimaging methods have revealed areas in nonprimary auditory cortex (in belt and parabelt regions) to be responsible for melody processing (Zatorre et al., 1994; Patterson et al., 2002; Brown & Martinez, 2007). In their fMRI study of melody processing, Patterson et al. (2002) presented two different types of melody, one in which 32 sequential IRN bursts produced a novel diatonic melody and one in which the IRN bursts produced a random note melody. Contrasting these two conditions with one in which there was a sequence of IRN bursts with a fixed pitch revealed activity within planum polare and superior temporal gyrus. Moreover, this activity was greater in the right hemisphere. The pronounced asymmetry emerged only for the effect of melody and was not present for the simple effect of pitch (defined by contrasting the fixed pitch sequence with a random noise condition). This finding is consistent with the hemispheric specialization hypothesis, which claims that the right hemisphere plays a dominant role in coding small and precise changes in frequency (pitch) over relatively long temporal durations (review: Zatorre et al., 2002).

The concept of a spatially segregated hierarchy of pitch coding has been proposed to explain the results presented (Patterson et al., 2002; Zatorre et al., 2002). At the first stage (possibly subcortical), temporal regularity is extracted from separate frequency channels of the incoming signal, while at the second stage (possibly lateral Heschl's gyrus) this temporal pattern information is integrated across frequency channels to code pitch. Higher-level processes such as pitch tracking and melody extraction occur at the third stage, especially in distributed regions of the right superior temporal gyrus and prefrontal cortex (Zatorre et al., 1994).

7.7 Summary

One possible overarching perspective of auditory cortex is a modular one in which sound recognition proceeds through several anatomically discrete and functionally specialized cortical areas culminating in higher centers where perceptual discriminations and other behaviorally relevant judgments are performed. Ventral and dorsal projections from lateral belt and parabelt regions to discrete regions of prefrontal cortex (as shown in Fig. 7.1A) are certainly not inconsistent with such a hierarchy model (Romanski et al., 1999). The neuroimaging results presented in this chapter show that a wide range of sounds from pure tones, through harmonic complex tones, modulated signals and pitches stimulate activity within primary and nonprimary regions of human auditory cortex. However, these data do not provide any strong sense in which key functional roles can be ascribed to the different anatomical regions illustrated in Figure 7.1 and are thus rather difficult to reconcile with the modular framework.

King and Nelken (2009) have recently proposed an alternative organizing principle within auditory cortex that goes beyond that of simple feature detection. They posit that primary auditory cortex (A1) sits at a higher level of processing than primary visual cortex (V1) and may be responsible for combining sound components across frequency and over time to generate interpretations of the auditory scene. Consequently, they argue that naturalistic stimuli are perhaps better suited for identifying the emergent properties within auditory cortex than well controlled synthesized signals. According to this argument, auditory cortical neurons are most sensitive to sounds that contain behaviorally relevant spectrotemporal patterns defined by a combination of different stimulus features. Such stimuli could be seen as "auditory objects" and this topic is discussed in more detail by Griffiths, Micheyl, and Overath (Chapter 8).

A slightly different perspective is that representations of simple sound features *are* topographically organized, but they are spatially distributed across the surface of auditory cortex. In the visual system, a body of evidence is beginning to demonstrate how cortical representations that were previously absent in the data might in reality be present (Grill-Spector et al., 2006; Logothetis, 2008). Clever experimental methodology and sophisticated data analysis are two key factors that may reveal organizations that might have previously been obscured. In terms of methodology, high-resolution imaging and fMRI adaptation designs are two examples that have been applied in the auditory domain. For example, Formisano et al. (2003) used a combination of ultra-high-field (7 Tesla) and surface coil fMRI to achieve a fine-grained spatial resolution (1.20 × 1.48 × 2.00 mm). High-resolution fMRI detected activity on a much finer spatial scale than had been reported hitherto, enabling mirror-symmetric frequency gradients on Heschl's gyrus to be measured systematically in each individual listener. fMRI adaptation designs are particularly recommended for investigating the functional properties of a brain region that has spatially overlapping or close neural populations that encode different stimulus categories (Grill-Spector et al., 2006). They are sensitive to differential fMRI responses within a region. The methods take advantage of the observation that the BOLD response decreases with repeated presentation of the same stimuli. In the auditory domain, fMRI adaptation studies have so far concerned the representation of perceptual categories such as speech sounds (Ahveninen et al., 2006; Leaver & Rauschecker, 2010), animal vocalizations (Altmann et al., 2007), and voice (Leaver & Rauschecker, 2010) instead of basic sound features such as pitch. For example, Altmann et al. (2007) reported that the response amplitude across the left superior temporal gyrus was significantly weaker for trials in which the same animal vocalization was repeated compared to trials in which the two animal vocalizations were different, thus indicating a selective representation of this sound category in left nonprimary auditory cortex.

The second key to discovering new principles of organization is to use clever analysis to maximize the potential afforded by clever design. Phase-encoded stimulus mapping and multivoxel pattern analysis are two examples that have been applied in the auditory domain. Unlike conventional pairwise contrast analysis, phase-encoded mapping compares the responses to a set of stimuli and estimates the most effective stimulus. For example, Talavage et al. (2004) were able to identify multiple tonotopic gradients systematically in individual listeners by mapping areas of auditory cortex that showed a progressive linear change in the frequency of maximal sensitivity. Another approach is to take into account the full spatial pattern of brain activity by applying a classification algorithm to decode what patterns are present across the cortical surface. Compared with univariate analysis, the particular strength of multivoxel pattern analysis is in revealing the representation of different perceptual categories within a single region of activity, often using discriminative responses that are weak but consistent across different sound examples. For example, using this method it has been shown that four sound categories evoke distinctive patterns of activity across the superior temporal gyrus (Staeren et al., 2009). A distributed cortical coding of sound properties could explain why several auditory regions have been implicated in the processing of many different auditory attributes. It is even possible that auditory cortical regions encoding relatively basic attributes of sounds (such as pitch) and higher level properties (such as category) are not mutually exclusive. Much more is known about basic sound processing in human auditory cortex than a decade or so ago. With recent interest in the application of novel approaches to fMRI design and analysis, there is every reason to be optimistic for the future.

Acknowledgments The second author received support for her work from the Medical Research Council through a Ph.D. studentship that was awarded from the MRC Institute of Hearing Research.

References

Ahveninen, J., Jääskeläinen, I. P., Raij, T., Bonmassar, G., Devore, S., Hämäläinen, M., et al. (2006). Task-modulated “what” and “where” pathway in human auditory cortex. *Proceedings of the National Academy of Sciences of the USA*, 103, 14608–14613.

Altmann, C. F., Doehrmann, O., & Kaiser, J. (2007). Selectivity for animal vocalizations in the human auditory cortex. *Cerebral Cortex*, 17, 2601–2608.

Bendor, D., & Wang, X. (2005). The neural representation of pitch in primate auditory cortex. *Nature*, 436, 1161–1165.

Brechmann, A., Baumgart, F., & Scheich, H. (2002). Sound-level-dependent representation of frequency modulations in human auditory cortex: A low-noise fMRI study. *Journal of Neurophysiology*, 87, 423–433.

Brosch, M., & Schreiner, C. E. (1999). Correlations between neural discharges are related to receptive field properties in cat primary auditory cortex. *European Journal of Neuroscience*, 11(10), 3517–3530.

Brown, S., & Martinez, M. J. (2007). Activation of premotor vocal areas during musical discrimination. *Brain and Cognition*, 63, 59–69.

Carlyon, R. P., Demany, L., & Semal, C. (1992). Detection of across-frequency differences in fundamental frequency. *Journal of the Acoustical Society of America*, 91, 279–292.

Chait, M., Poeppel, D., & Simon, J. Z. (2006). Neural response correlates of detection of monaurally and binaurally created pitches in humans. *Cerebral Cortex*, 16, 835–848.

Chatterjee, M., & Zwislocki, J. J. (1998). Cochlear mechanisms of frequency and intensity coding. II. Dynamic range and the code for loudness. *Hearing Research*, 124, 170–181.

Dean, I., Robinson, B. L., Harper, N. S., & McAlpine, D. (2008). Rapid neural adaptation to sound level statistics. *Journal of Neuroscience*, 28, 6430–6438.

de Cheveigné, A. (2005). Pitch perception models. In C. J. Plack, A., Oxenham, Fay, R. R., & Popper, A. N. (Eds.), *Pitch: Neural coding and perception* (pp. 169–233).New York: Springer-Verlag.

Formisano, E., Kim, D. S., Di Salle, F., van de Moortele, P. F., Ugurbil, K., & Goebel, R. (2003). Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron*, 40(4), 859–869.

Garcia, D., Hall, D. A., & Plack, C. J. (2010). The effect of stimulus context on pitch representations in the human auditory cortex. *NeuroImage*, 51, 808–816.

Giraud, A.-L., Lorenzi, C., Ashburner, J., Wable, J., Johnsrude, I., Frackowiak, R., & Kleinschmidt, A. (2000). Representation of the temporal envelope of sounds in the human brain. *Journal of Neurophysiology*, 84, 1588–1598.

Griffiths, T. D., Büchel, C., Frackowiak, R. S. J., & Patterson, R. D. (1998). Analysis of temporal structure in sound by the human brain. *Nature Neuroscience*, 1, 422–427.

Griffiths, T. D., Kumar, S., Sedley, W., Nourski, K. V., Kawasaki, H., Oya, H., et al. (2010). Direct recordings of pitch responses from human auditory cortex. *Current Biology*, 20, 1128–1132.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, 10, 14–23.

Hackett, T. A. (2003). The comparative anatomy of the primate auditory cortex. In A. A. Ghazanfar (Ed.), *Primate audition: Ethology and neurobiology* (pp.199–225).Boca Raton, FL: CRC Press.

Hall, D. A., & Plack, C. J. (2009). Pitch processing sites in the human auditory brain. *Cerebral Cortex*, 19, 576–585.

Hall, D. A., Haggard, M. P., Summerfield, A. Q., Akeroyd, M. A., Palmer, A. R., & Bowtell, R. W. (2001). Functional magnetic resonance imaging measurements of sound-level encoding in the absence of background scanner noise. *Journal of the Acoustical Society of America*, 109, 1559–1570.

Hall, D. A., Johnsrude, I. S., Haggard, M. P., Palmer, A. R., Akeroyd, M. A., & Summerfield, A. Q. (2002). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, 12, 140–149.

Hall, D. A., Barrett, D. J. K., Akeroyd, M. A., & Summerfield, A. Q. (2005). Cortical representations of temporal structure in sound. *Journal of Neurophysiology*, 94, 3181–3191.

Harms, M. P., & Melcher, J. R. (2002). Sound repetition rate in the human auditory pathway: Representations in the waveshape and amplitude of fMRI activation. *Journal of Neurophysiology*, 88, 1433–1450.

Harms, M. P., Guinan, J. J., Sigalovsky, I. S., & Melcher, J. R. (2005). Short-term sound temporal envelope characteristics determine multisecond time patterns of activity in human auditory cortex as shown by fMRI. *Journal of Neurophysiology*, 93, 210–222.

Hart, H. C., Palmer, A. R., & Hall, D. A. (2002). Heschl's gyrus is more sensitive to tone level than non-primary auditory cortex. *Hearing Research*, 171, 177–190.

Hart, H. C., Palmer, A. R., & Hall, D. A. (2003a). Amplitude and frequency-modulated stimuli activate common regions of human auditory cortex. *Cerebral Cortex*, 13, 773–781.

Hart, H. C., Palmer, A. R., & Hall, D. A. (2003b). The sound-level-dependent growth in the extent of fMRI activation in Heschl's gyrus is different for low- and high-frequency tones. *Hearing Research*, 179, 104–112.

Hart, H. C., Palmer, A. R., & Hall, D. A. (2004). Different areas of human non-primary auditory cortex are activated by sounds with spatial and nonspatial properties. *Human Brain Mapping*, 21, 178–190.

Houtsma, A. J. M., & Smurzynski, J. (1990). Pitch identification and discrimination for complex tones with many harmonics. *Journal of the Acoustical Society of America*, 87, 304–310.

Jäncke, L., Shah, N. J., Posse, S., Grosse-Ryuken, M., & Muller-Gartner, H.-W. (1998). Intensity coding of auditory stimuli: An fMRI study. *Neuropsychologia*, 36, 875–883.

Joris, P. X., & Yin, T. C. T. (1992). Responses to amplitude-modulated tones in the auditory nerve of the cat. *Journal of the Acoustical Society of America*, 91, 215–232.

Kaas, J. H., & Hackett, T. A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Sciences of the USA*, 97, 11793–11799.

King, A. J., & Nelken, I. (2009). Unraveling the principles of auditory cortical processing: Can we learn from the visual system? *Nature Neuroscience*, 12, 698–701.

Krumbholz, K., Patterson, R. D., Seither-Preisler, A., Lammertmann, C., & Lütkenhöner, B. (2003). Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. *Cerebral Cortex*, 13, 765–772.

Langers, D. R. M., Backes, W. H., & van Dijk, P. (2003). Spectrotemporal features of the auditory cortex: The activation in response to dynamic ripples. *NeuroImage*, 20, 265–275.

Langers, D. R. M., Backes, W. H., & van Dijk, P. (2007a). Representation of lateralization and tonotopy in primary versus secondary human auditory cortex. *NeuroImage*, 34, 264–273.

Langers, D. R. M., van Dijk, P., Schoenmaker, E. S., & Backes, W. H. (2007b). fMRI activation in relation to sound intensity and loudness. *NeuroImage*, 35, 709–718.

Lasota, K. J., Ulmer, J. L., Firszt, J. B., Biswal, B. B., Daniels, D. L., & Prost, R. W. (2003). Intensity-dependent activation of the primary auditory cortex in functional magnetic resonance imaging. *Journal of Computer Assisted Tomography*, 27, 213–218.

Leaver, A. M., & Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: Effects of acoustic features and auditory object category. *Journal of Neuroscience*, 30, 7604–7612.

Liang, L., Lu, T., & Wang, X. (2002). Neural representations of sinusoidal amplitude and frequency modulations in the primary auditory cortex of awake primates. *Journal of Neurophysiology*, 87, 2237–2261.

Liberman, M. C. (1978). Auditory-nerve responses from cats raised in a low-noise chamber. *Journal of the Acoustical Society of America*, 63, 442–455.

Lockwood, A. H., Salvi, R. J., Coad, M. L., Arnold, S. A., Wack, D. A., Murphy, B. W., & Burkard, R. F. (1999). The functional anatomy of the normal human auditory system: Responses to 0.5 and 4.0 kHz tones at varied intensities. *Cerebral Cortex*, 9, 65–76.

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, 453, 869–878.

Lu, T., Liang, L., & Wang, X. (2001). Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. *Nature Neuroscience*, 4, 1131–1138.

Luo, H., Wang, Y., Poeppel, D., & Simon, J. Z. (2007). Concurrent encoding of frequency and amplitude modulation in human auditory cortex: Encoding transition. *Journal of Neurophysiology*, 98, 3473–3485.

Manunta, Y., & Edeline, J. M. (1998). Effects of noradrenaline on rate-level function of auditory cortex neurons: Is there a "gating" effect of noradrenaline? *Experimental Brain Research*, 118, 361–372.

Mohr, C. M., King, W. M., Freeman, A. J., Briggs, R. W., & Leonard, C. M. (1999). Influence of speech stimuli intensity on the activation of auditory cortex investigated with functional magnetic resonance imaging. *Journal of the Acoustical Society of America*, 105, 2738–2745.

Moore, B. C. J. (2004). *An introduction to the psychology of hearing*, 5th ed. London: Elsevier.

Moore, B. C. J., Glasberg, B. R., & Baer, T. (1997). A model for the prediction of thresholds, loudness, and partial loudness. *Advances in Engineering Software*, 45, 224–240.

Morel, A., Garraghty, P. E., & Kaas, J. H. (1993). Tonotopic organization, architectonic fields, and connections of auditory cortex in macaque monkeys. *Journal of Comparative Neurology*, 335, 437–459.

Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T., & Zilles, K. (2001). Human primary auditory cortex: Cytoarchitectonic subdivisions and mapping into a spatial reference system. *NeuroImage*, 13, 684–701.

Pantev, C., Hoke, M., Lehnertz, K., Lütkenhöner, B., Anogianakis, G., & Wittkowski, W. (1988). Tonotopic organization of the human auditory cortex revealed by transient auditory evoked magnetic fields. *Electroencephalography and Clinical Neurophysiology*, 69, 160–170.

Pantev, C., Hoke, M., Lütkenhöner, B., & Lehnertz, K. (1989). Tonotopic organization of the auditory cortex: Pitch versus frequency representation. *Science*, 246, 486–488.

Pantev, C., Bertrand, O., Eulitz, C., Verkindt, C., Hampson, S., Schuierer, G., & Elbert, T. (1995). Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings. *Electroencephalography and Clinical Neurophysiology*, 94, 26–40.

Pantev, C., Elbert, T., Ross, B., Eulitz, C., & Terhardt, E. (1996). Binaural fusion and the representation of virtual pitch in the human auditory cortex. *Hearing Research*, 100, 164–170.

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, 36, 767–776.

Penagos, H., Melcher, J. R., & Oxenham, A. J. (2004). A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *Journal of Neuroscience*, 24, 6810–6815.

Phillips, D. P., Orman, S. S., Musicant, A. D., & Wilson, G. F. (1985). Neurons in the cat's primary auditory cortex distinguished by their responses to tones and wide-spectrum noise. *Hearing Research*, 18, 87–102.

Phillips, D. P., Semple, M. N., Calford, M. B., & Kitzes, L. M. (1994). Level-dependent representation of stimulus frequency in cat primary auditory cortex. *Experimental Brain Research*, 102, 210–226.

Rauschecker, J. P., Tian, B., & Hauser, M. (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. *Science*, 268, 111–114.

Recanzone, G. H., Schreiner, C. E., Sutter, M. L., Beitel, R. E., & Merzenich, M. M. (1999). Functional organization of spectral receptive fields in the primary auditory cortex of the owl monkey. *The Journal of Comparative Neurology*, 415, 460–481.

Rivier, F., & Clarke, S. (1997). Cytochrome oxidase, acetylcholinesterase, and NADPH-diaphorase staining in human supratemporal and insular cortex: Evidence for multiple auditory areas. *NeuroImage*, 6, 288–304.

Romanski, L. M., Bates, J. F., & Goldman-Rakic, P. S. (1999). Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. *The Journal of Comparative Neurology*, 403, 141–157.

Saberi, K., & Hafter, E. R. (1995). A common neural code for frequency- and amplitude-modulated sounds. *Nature*, 374, 537–539.

Schönwiesner, M., & Zatorre, R. J. (2008). Depth electrode recordings show double dissociation between pitch processing in lateral Heschl's gyrus and sound onset processing in medial Heschl's gyrus. *Experimental Brain Research*, 187, 97–105.

Schönwiesner, M., von Cramon, D. Y., & Rübsamen R. (2002). Is it tonotopy after all? *NeuroImage*, 17, 1141–1161.

Sigalovsky, I. S., & Melcher, J. R. (2006). Effects of sound level on fMRI activation in human brainstem, thalamic and cortical centers. *Hearing Research*, 215, 67–76.

Staeren, N., Renvall, H., De Martino, F., Goebel, R., & Formisano, E. (2009). Sound categories are represented as distributed patterns in the human auditory cortex. *Current Biology*, 19, 498–502.

Stufflebeam, S. M., Poeppel, D., Rowley, H. A., & Roberts, T. P. L. (1998). Peri-threshold encoding of stimulus frequency and intensity in the M100 latency. *NeuroReport*, 9, 91–94.

Talavage, T. M., Ledden, P. J., Benson, R. R., Rosen, B. R., & Melcher, J. R. (2000). Frequency-dependent responses exhibited by multiple regions in human auditory cortex. *Hearing Research*, 150, 225–244.

Talavage, T. M., Sereno, M. I., Melcher, J. R., Ledden, P. J., Rosen, B. R., & Dale, A. M. (2004). Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. *Journal of Neurophysiology*, 91, 1282–1296.

Upadhyay, J., Ducros, M., Knaus, T. A., Lindgren, K. A., Silver, A., Tager-Flusberg, H., & Kim D.-S. (2007). Function and connectivity in human primary auditory cortex: A combined fMRI and DTI study at 3 Tesla. *Cerebral Cortex*, 17(10), 2420–2432.

Viemeister, N. F., & Bacon, S. P. (1988). Intensity discrimination, increment detection, and magnitude estimation for 1-kHz tones. *Journal of the Acoustical Society of America*, 84, 172–178.

Wallace, M. N., Johnston, P. W., & Palmer, A. R. (2002). Histochemical identification of cortical areas in the auditory region of the human brain. *Experimental Brain Research*, 143, 499–508.

Wang, X., & Sachs, M. B. (1995).Transformation of temporal discharge patterns in a ventral cochlear nucleus stellate cell model: Implications for physiological mechanisms. *Journal of Neurophysiology*, 73, 1600–1616.

Wessinger, C. M., Buonocore, M. H., Kussmaul, C. L., & Mangun, G. R. (1997). Tonotopy in human auditory cortex examined with functional magnetic resonance imaging. *Human Brain Mapping*, 5, 18–25.

Wessinger, C. M., VanMeter, J., Tian, B., Van Lare, J., Pekar, J., & Rauschecker, J. P. (2001). Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *Journal of Cognitive Neuroscience*, 13, 1–7.

Yost, W. (1996). Pitch of iterated rippled noise. *Journal of the Acoustical Society of America*, 100, 511–518.

Yost, W. A., Patterson, R., & Sheft, S. (1996). A time domain description for the pitch strength of iterated rippled noise. *Journal of the Acoustical Society of America*, 99, 1066–1078.

Zatorre, R. J., Evans, A. C., & Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *Journal of Neuroscience*, 14, 1908–1919.

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences*, 6, 37–46.

Chapter 8
Auditory Object Analysis

Timothy D. Griffiths, Christophe Micheyl, and Tobias Overath

8.1 The Problem

The concept of what constitutes an auditory object is controversial (Kubovy & Van Valkenburg, 2001; Griffiths & Warren, 2004; Nelken, 2004). It is more difficult to examine the sound pressure waveform that enters the cochlea and "see" different objects in the same way that we "see" objects in the visual input to the retina. However, in both the auditory system and the visual system, objects can be understood in terms of the "images" they produce during the processing of sense data. The idea that objects are mental events that result from the creation of images from sense data goes back to Kant (1929). Visual images, representations in the visual brain corresponding to objects, can be understood as having two spatial dimensions. These arrays of neural activity preserve spatial relationships from the retina to the cortex. Auditory images, which may be thought of as representations in the auditory brain that correspond to objects, can also be considered in terms of the dimensions of the signal processed by the cochlea. The critical step is to consider that signal, not as a sound pressure waveform, but as a signal with dimensions of frequency, represented across the receptor array, and time. It is argued here that analysis of images

T.D. Griffiths (✉)
Institute of Neuroscience, Newcastle University Medical School,
Framlington Place, Newcastle upon Tyne, NE2 4HH, UK
e-mail: t.d.griffiths@newcastle.ac.uk

C. Micheyl
Auditory Perception and Cognition Laboratory, Department of Psychology,
University of Minnesota, N628 Elliot Hall, 75 East River Road, Minneapolis, MN 55455, USA
e-mail: cmicheyl@umn.edu

T. Overath
Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, USA
e-mail: t.overath@nyu.edu

D. Poeppel et al. (eds.), *The Human Auditory Cortex*, Springer Handbook
of Auditory Research 43, DOI 10.1007/978-1-4614-2314-0_8,

with dimensions related to frequency and time is a helpful way of considering auditory analysis in the pathway from cochlea to cortex. Many syntheses of this process are based on derivations of frequency and time such as spectral ripple density related to frequency and amplitude modulation (Chi et al., 2005) or forms of autocorrelation (Patterson, 2000) related to time. But the principle of objects existing in a space with dimensions of frequency and time remains the same, and if one accepts the existence of images with a temporal dimension, then the concepts of auditory objects and auditory images can be considered in a comparable way to the visual system. The idea was first proposed by Kubovy and Van Valkenburg (2001), who suggested that auditory objects can be considered as existence regions within frequency–time space that have borders with the rest of the sound scene.

A second source of controversy, which also happens to apply to visual analysis, is the cognitive level to which the concept of auditory object analysis should be extended. Consider the situation in which you hear someone making the vowel sound /u/ at a pitch of 220 Hz and intensity of 60 dB on the left side of the room. That situation requires sensory analysis of the spectrotemporal structure of the sound. It also requires categorical perception to allow the sound to be distinguished from other sounds. Sounds from which it has to be distinguished might be from another class (e.g., a telephone ringing at the same pitch, intensity, and location) or the same class (e.g., another person making the vowel sound /u/ at a different pitch, intensity, or spatial location). We can appreciate that we are listening to the same type of sound if we hear it at 150 Hz or 80 dB or on the right side of the room. We can appreciate that similarity, even if we do not speak a relevant language to allow us to recognize or name the vowel. At another level of analysis the sound must enter a form of echoic memory store (to allow comparison with sounds that might immediately follow it) and might enter an anterograde memory store that allows comparison with sounds heard over days or weeks. At a further level of analysis we might call the sound a voice, or my voice, the vowel "u," or (if we have absolute pitch) "A3." The term object analysis might therefore be applied to (1) the perception of a coherent whole, the essence of which can be perceived even when cues such as pitch or intensity are changed; (2) categorical analysis; (3) encoding into working memory; (4) encoding into anterograde memory; or (5) association with a label during semantic analysis. The term object analysis is used for the first stage here and it is emphasized here, but a number of workers would argue for an obligatory requirement for auditory objects, or objects in general, to have an associated label. The key point, however, is that there are a number of aspects of analysis of the coherent whole that are defined as object analysis here, and that even this first stage requires considerable computational work to derive a representation of particular sounds that is independent of basic cues such as pitch or intensity. Clinically, the distinction of presemantic and semantic processing stages of object analysis is relevant to the existence of apperceptive and associative forms of auditory agnosia, respectively (Griffiths et al., 2009a).

A third controversial aspect of auditory object analysis is whether the concept should be applied to particular individual sounds that can be distinguished from others as argued above or to *sequences* of sounds that are grouped: auditory streams. Bregman (1990) explicitly rejects the concept of auditory object in favor of the auditory stream, a sequence of grouped sounds, as the fundamental unit of auditory perception. Others have equated auditory objects with streams (Shamma, 2008). There is a problem, however, with considering streams as auditory objects corresponding to a single percept derived from analysis over longer periods of time. Streams are sequences of sounds that are grouped by perceptual properties such as pitch, timbre, or position (Moore & Gockel, 2002), and these perceptual properties all have complex relationships to the acoustic structure. This makes a description of the stream as the most fundamental unit of analysis leading to perception problematic, when the stream itself comprises elements that are perceived individually. Whether streams are regarded as objects, streams of objects, or something else, it is nevertheless an important level of perceptual organization that is considered here.

At a more generic level, auditory objects can be conceptualized in information theoretic terms, such that a given auditory object is characterized through its probabilistic higher-order statistical properties; in turn, boundaries between auditory objects are indicated by transitions in these statistical regularities (Kubovy & Van Valkenburg, 2001). That is, at a general descriptive level, auditory objects are defined in terms of their distinct statistical signal characteristics, which simultaneously distinguish them from other auditory objects (and possibly other object classes). Statistical regularities thus provide important information for auditory scene analysis, as they allow the perceptual organization of the acoustic environment including figure–ground segregation.

This chapter considers brain bases for aspects of object analysis within a serial framework. It starts with segregation and simultaneous grouping mechanisms that allow the perception of coherent “whole” percepts (objects) that can be associated with properties including pitch and timbre, before a consideration of sequential grouping and higher-level analysis. Although early simultaneous grouping may have a subcortical contribution (Pressnitzer et al., 2001), here it is argued that in humans the analysis of auditory objects is critically dependent on processing in auditory cortex, with some evidence for a hierarchy of bases for analysis that parallels the perceptual hierarchy. For the purpose of this chapter, there is no need for any of the stages considered here to be regarded as “the” essence of object analysis: rather, the chapter attempts a systematic approach to object analysis that addresses the major levels that might be relevant. Overall, the process might be regarded as the abstraction and organization of things from the acoustic world as an extension of the sensory representation of frequency-time structure considered by Hall and Barker, Chapter 7. Critically, this chapter moves from the level of sensory representation to perception.

As noted earlier, this chapter considers the main anatomical substrate for the analysis of auditory objects to be auditory cortex. Relevant behavioral aspects in the corresponding subsections are also briefly mentioned; this is not, however, intended to be a complete synthesis of the behavioral data on auditory objects. This chapter focuses on human studies but mentions animal work that might illuminate neuronal mechanisms also relevant to humans.

8.2 Simultaneous Grouping and Object Segregation

A critical aspect of object analysis is the lumping together of elements of the sound scene that constitute the object as opposed to elements that do not. The idea that auditory objects might generally be defined as regions in auditory-time space with borders (Kubovy & Van Valkenburg, 2001) was explored in a study in which the timbre of the object and the salience of the border were systematically manipulated (Overath et al., 2010). Figure 8.1 shows the synthetic stimulus used to investigate this, known as an acoustic "texture." The stimulus is based on linear ramps in frequency-time space with random timing and frequency of onset, as well as given excursions. The textures are associated with a perceptual quality that can be systematically manipulated by changing the coherence of the ramps: the percentage of ramps with a particular excursion. The appreciation of this perceptual quality requires participants to abstract ensemble properties of frequency–time space within probabilistic constraints and irrespective of stochastic variation—stimuli with the same coherence sound like the same thing or object despite differences in their detailed structure.

The manipulation of adjacent textures shown in Figure 8.1 has two effects. First, boundaries between different adjacent textures are defined by changes in the coherence across that boundary, where the salience of the boundary depends on the difference in coherence on either side of the boundary as opposed to the absolute coherence values. The stimulus has no physical "edge" like an onset from silence and the spectral density of the stimulus over time is kept constant. Second, the absolute coherence of the different regions is associated with a particular perceptual quality or timbre. Figure 8.2 shows a functional magnetic resonance imaging (fMRI; see Talavage, Johnsrude, and Gonzalez Castillo, Chapter 6) study in which these two aspects—object segregation and representation—were manipulated orthogonally.

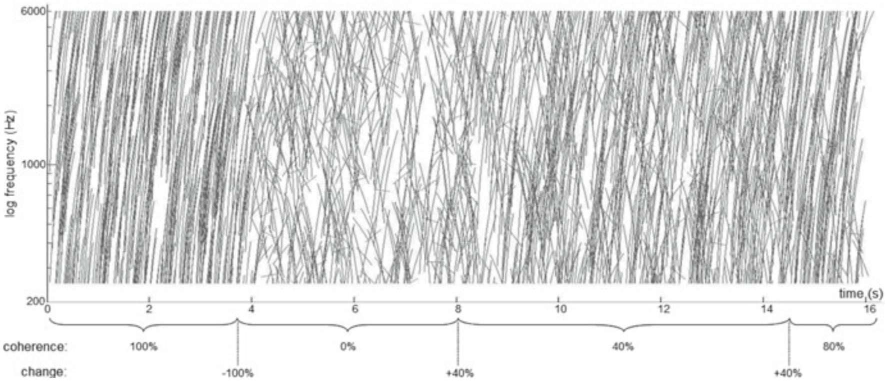


Fig. 8.1 Stimulus for investigation of object-segregation mechanisms (Overath et al., 2010). Example of a block of sound with four spectrotemporal coherence segments showing absolute coherence values for each segment and the corresponding change in coherence between the segments. (Adapted from Overath et al., 2010.)

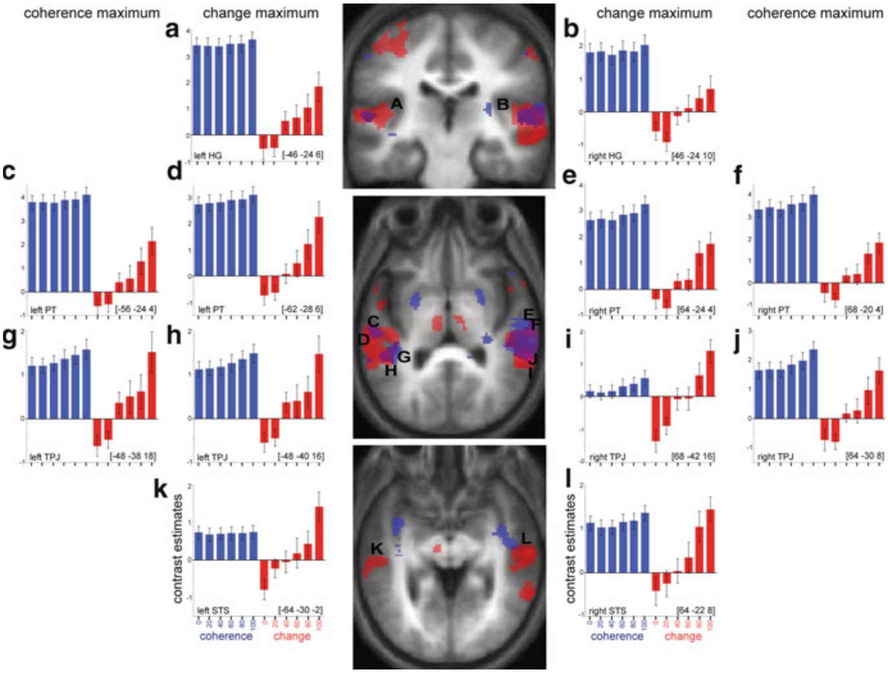


Fig. 8.2 Brain mechanisms for object-segregation mechanisms. Areas showing an increased hemodynamic response as a function of increasing absolute coherence (blue) and increasing change in coherence (red). Results are rendered on coronal (*y* = –24, top) and tilted (pitch = –0.5, middle [superior temporal plane] and bottom [STS]) sections of participants' normalized average structural scans. The bar charts show the mean contrast estimates (± SEM) in a sphere with 10 mm radius around the local maximum corresponding to the six levels of absolute coherence (blue) and the six levels of change in coherence (red). Change in coherence levels are pooled across "positive" and "negative" changes so as to show the main effect of change in coherence magnitude. The charts nearest the brain show the mean response in the sphere around the local maxima for increasing change in coherence; those at the sides show the mean response in the sphere around local maxima for increasing absolute coherence. Note that the placement of the identifying letter in the brain sections only approximate the precise stereotactic (*x*, *y*, *z*) coordinates at the bottom corner of each chart because no single planar section can contain all the local maxima simultaneously. (Adapted from Overath et al., 2010.)

The data show an effect of boundary salience in a number of areas including primary auditory cortex in the medial part of Heschl's gyrus (HG), while the absolute coherence was represented in auditory cortex beyond HG. The data are therefore consistent with the existence of an early mechanism for the segregation of objects that is distinct from the associated perceived timbre. These data do not allow any comment on the order of analysis but a reasonable hypothesis is the existence of a serial mechanism in which segregation precedes timbre perception. Another way to consider the data is in terms of the size of spectrotemporal regions that must be analyzed to perceive either a change in object or a perceptual quality. Object change might be achieved by the perception of a local rule change while analysis of the perceptual properties of the object within the boundary necessarily involves analysis of larger regions of frequency-time space. The data are therefore also consistent with the idea that more extended segments of spectrotemporal space are analyzed in areas further from primary cortex. This idea is supported by time-domain studies showing longer windows of analysis as one moves from primary cortex (Boemio et al., 2005; Overath et al., 2008).

In addition to boundary recognition, a number of other mechanisms for the simultaneous grouping of elements into objects have been established by behavioral experiments (Darwin & Carlyon, 1995; Ciocca, 2008). Simultaneous onset might be regarded as a particular type of vertical boundary in frequency-time space and is a strong grouping cue. Another strong grouping cue is the presence of harmonic relationships between frequency elements (harmonicity), where elements exist at frequencies that are integer multiples of a given fundamental frequency. Note that harmonicity is a cue that can allow the separation of objects that occur in overlapping regions of frequency-time space where there is no clear boundary, for example, when two speakers produce vowels at the same time. Common modulation of elements is a weaker grouping cue, as is the spatial location of a source.

Electroencephalography (EEG; see also Alain and Winkler, Chapter 4) and magnetoencephalography (MEG; see also Nagarajan, Gabriel, and Herman, Chapter 5) have also been used to investigate simultaneous grouping mechanisms in auditory cortex. Mistuning of one harmonic of a harmonic complex to disrupt harmonicity is accompanied by an EEG evoked response at about 30 ms (Pa) attributed to primary auditory cortex (Dyson & Alain, 2004) and also a later response with a latency of approximately 150 ms that has been called an object related negativity (ORN; Alain & Izenberg, 2003; see also Alain and Winkler, Chapter 4). Later responses after 400 ms are also described that, unlike the earlier responses, are strongly influenced by attention. Recent work based on MEG (Lipp et al., 2010) has demonstrated a magnetic equivalent of the ORN (ORNm) that is sensitive to both harmonicity and common onset but not attention. The ORNm can be argued to be a correlate of a generic simultaneous grouping process (that can operate on different cues) occurring in auditory cortex.

These studies suggest a critical role of human auditory cortex in simultaneous grouping during object analysis beyond the simple representation of stimulus cues such as frequency and amplitude. The response to texture boundaries (Overath et al., 2010) suggests early segregation or change detection mechanisms in primary cortex, while the ORNm responses (Alain & Izenberg, 2003; Lipp et al., 2010) have longer latencies than the middle latency responses attributed to primary cortex.

8.2.1 Object Features Based on Segregated and Grouped Elements

Simultaneous grouping is a basis for the abstraction of patterns that are associated with perceived features. In the case of grouping by harmonicity, the coherent whole that is perceived is associated with pitch. Considered in this way, pitch is related to ensemble sound properties considered in the frequency domain, and the earliest theories of pitch perception were indeed based on the frequency composition of the stimulus (von Helmholtz, 1885). Most modern pitch theories also emphasize stimulus properties in the time domain, especially temporal regularity (review: de Cheveigné, 2005). The relationship between stimulus properties in the frequency or time domain and the neural correlates of pitch is therefore necessarily a complex one: a simple relationship between auditory cortical areas (defined as areas containing separate gradients corresponding to stimulus frequency) and any representation of perceived pitch would not necessarily be expected, a priori.

Marmoset (*Callithrix jacchus*) recordings from single auditory neurons (Bendor & Wang, 2005) have identified a cortical subarea abutting primary cortex in which the responses were "tuned" to the pitch of the stimulus rather than its frequency composition. In contrast, recordings from single units in ferrets (*Mustela putorius*) (Bizley et al., 2009) showed an effect of pitch value on the responses of single neurons in multiple cortical areas, although based on a less strict criterion for a pitch-responsive neuron. In humans, measurements from neural ensembles with MEG (Krumbholz et al., 2003) have shown responses to the transition between a noise and a regular-interval-noise associated with pitch in primary auditory cortex. Measurements from neural ensembles based on the fMRI blood oxygenation level–dependent (BOLD) response (Patterson et al., 2002) have shown responses to regular noise in lateral HG in nonprimary cortex. Human fMRI studies based on a broader range of stimuli including a type of spatial pitch (Huggins pitch) have not consistently demonstrated involvement of HG (Hall & Plack, 2009; Puschmann et al., 2009). There are therefore a number of unresolved issues related to the representation of pitch, including: (1) whether pitch perception can ever be adequately characterized by the properties of single neurons or whether it will require characterization based on ensemble properties of neurons, as in the case of spatial location (Middlebrooks et al., 1994; Miller & Recanzone, 2009); (2) which of the aforementioned responses might relate to stimulus properties (especially regularity) and which reflect the perceived pitch; (3) the extent to which a common mechanism might be found across all species, or even just primates (>200 species); and (4) the way in which different areas might show effective connectivity to produce a specific pitch system. The data point to an important role of auditory cortex in representing the object property of pitch, but a great deal more work is required in this area.

Grouped elements also have perceptual properties distinct from pitch. Timbre, the characteristic that determines whether two sounds with the same pitch and loudness are dissimilar ("Acoustical terminology," 1960), is a multidimensional property that also has a complex relationship to acoustic structure (for a more detailed discussion, see Griffiths et al., 2009b). Some aspects of timbre, such as "brightness," are better explained by the spectral characteristics of the stimulus and other aspects, such as "attack," by the temporal characteristics (McAdams et al., 1995). Human fMRI and modeling experiments in which the spectral dimension of timbre was manipulated (Warren et al., 2005; Kumar et al., 2007) suggest a hierarchal model for this aspect of timbre analysis. In the model (Kumar et al., 2007), object features are abstracted in HG before further analysis in the right planum temporale (PT) posterior to HG and then the right anterior superior temporal sulcus. The model is consistent with the early grouping mechanisms in primary cortex suggested above followed by symbolic processing in more distributed temporal lobe regions.

8.3 Sequential Grouping

Sequential auditory grouping processes are frequently referred to as auditory "streaming" (Bregman, 1990). The first and main part of this section is concerned with auditory streaming and its neural basis in and beyond auditory cortex. Subsequently, other aspects of the perception and cortical processing of sound sequences are considered. In particular, the neural correlates of the auditory continuity illusion in auditory cortex are discussed, as well as neural correlates of the analysis of statistical regularities in sound sequences over time.

8.3.1 Auditory Streaming

In the past decade, a growing number of studies have investigated neural correlates of auditory streaming in auditory cortex, using techniques ranging from single- and multiunit recordings in nonhuman species, to EEG, MEG, and fMRI in humans. This chapter summarizes the main findings of these studies, focusing on humans. Because these findings cannot be understood without some knowledge of basic psychophysical facts concerning auditory streaming, it starts with a very brief overview of these facts. Note that, in the spirit of this book, the objective of this section is merely to provide an overview of an active topic of research, aimed primarily at non-experts readers. Additional information concerning the neural correlates of auditory streaming in both animals and humans can be found in other publications, including short overviews (Carlyon, 2004; Shamma & Micheyl, 2010), more detailed review articles (Micheyl et al., 2007; Snyder & Alain, 2007; Bee & Micheyl, 2008), and book chapters (Fay, 2008; Fishman & Steinschneider, 2010).

8.3.1.1 Basic Psychophysics of Auditory Streaming

Essential features of auditory streaming can be demonstrated using sound sequences, which are formed by presenting two sounds, A and B, in a repeating ABAB or ABA-ABA pattern (Fig. 8.1). Typically, the A and B sounds are pure tones with equal levels but different frequencies. It has been found that if the frequency separation (labeled Δf in Fig. 8.3A) is relatively small (say, 1 semitone, or approximately 6%) listeners hear the stimulus sequence as a single coherent "stream." In contrast, at large frequency separations, most listeners report hearing two separate streams of constant-pitch tones. The percept depends not only on the frequency separation between the A and B tones but also on the presentation rate: in general, fast

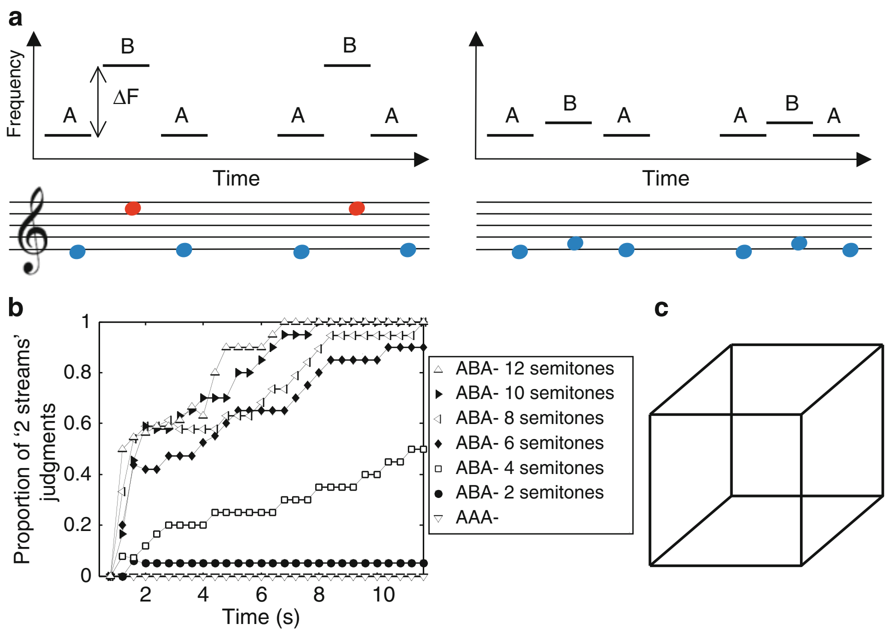


Fig. 8.3 Stimuli and percepts commonly used in studies of auditory streaming. (**a**) Schematic spectrograms of sound sequences used to study auditory streaming, and corresponding percepts. The stimulus sequences are formed by presenting two tones at different frequencies, A and B, in a repeating ABA-ABA… pattern, where the dash (-) stands for a silent gap. The dominant percept evoked by such sequences depends on the frequency separation (ΔF) between the A and B tones. When ΔF is relatively large, e.g., 7 semitones (left panel in **a**) the sequence is usually perceived as two separate "streams" of sounds: a low-pitch stream (shown in blue on the left musical partition), and a higher-pitch, lower-tempo stream (shown in red on the left musical partition). When ΔF is small, e.g., 1 semitone (right panel in **a**), the sequence is usually perceived as a single stream with a "galloping" rhythm. (**b**) The "buildup" of stream segregation. This figure illustrates how the proportion of trials for which a listener heard sequences such as those illustrated in **a** as "two streams" at a given time (relative to sequence onset). The different symbols correspond to different ΔF values, in semitones. As can be seen, except for the zero-ΔF (AAA) condition, the proportion of "two streams" judgments increases over time. The increase usually becomes more marked and more rapid as ΔF increases. (**c**) The Necker cube. When the transparent cube is looked at for a long-enough period of time, its orientation appears to "switch." Changes in auditory percept from "one stream" to "two streams" and vice versa during prolonged listening to sequences such as those illustrated in **a** may be thought of a an auditory counterpart of the percept reversals evoked by ambiguous figures such as the Necker cube

tone-presentation rates tend to promote segregation, whereas slow rates tend to promote integration (van Noorden, 1975).

Importantly, the percept evoked by these alternating tone sequences can fluctuate over time, even when the physical stimulus remains constant. The sequence is usually heard as a single stream at first; however, if the frequency separation is large enough, the percept usually switches to two separate streams after a few seconds of uninterrupted listening (Anstis & Saida, 1985). This delay varies stochastically across trials, even in a given individual (Pressnitzer & Hupé, 2006). Thus, when the proportion of trials on which the switch has occurred by time *t* (the time since sequence onset) is plotted as a function of *t*, it is usually found to increase gradually over the first 5–10 s (Carlyon et al., 2001) (Fig. 8.3B). This phenomenon has been dubbed the "buildup" of stream segregation. It has been suggested that the buildup depends critically upon attention, and does not occur (or at least, does not occur as much) if the listener's attention is directed away from the tone sequence, for example, toward the opposite ear, or toward a visual stimulus (Carlyon et al., 2001).

If listening continues beyond the first switch from one to two streams, the percept occasionally reverts back to that of a single stream. Upon more protracted listening, switches back and forth between one stream and two streams are experienced, at an average period of several seconds (Pressnitzer & Hupé, 2006). This phenomenon is reminiscent of bistability in visual perception; for instance, upon prolonged viewing, a Necker cube can be seen to switch back and forth between two orientations. Changes in percept that occur in the absence of concomitant changes in the physical stimulus have played (and continue to play) a key role in the identification of neural correlates of conscious visual experience (e.g., Logothetis & Schall, 1989). Thus, it is not surprising that the buildup, and the subsequent spontaneous switches between one-stream and two-stream percepts, have inspired several studies of the neural correlates of auditory streaming, both in humans (Cusack, 2005; Gutschalk et al., 2005; Snyder et al., 2006) and in other species (Micheyl et al., 2005b; Pressnitzer et al., 2008). These studies are discussed in a subsequent section.

If auditory streaming occurred only with pure tones, it would have little relevance to hearing in everyday life. However, the phenomenon has been observed with many other types of sounds, including harmonic complex tones, noises, and even synthetic vowels, as well as along dimensions other than frequency separation, including pitch determined by fundamental frequency (F0) or modulation rate, and timbre determined by spectral or temporal cues (for a review, see Moore & Gockel, 2002). It has been suggested that any physical difference that yields a sufficiently salient perceived difference between consecutive sounds can potentially be exploited by the auditory system to form separate streams (Moore & Gockel, 2002).

Another reason why auditory streaming may play an important role in hearing is that the way in which a sound sequence is organized perceptually appears to have a dramatic impact on the listener's ability to perceive detailed features of that sequence. For instance, listeners find it difficult to identify the temporal order of tones across different streams (Bregman & Campbell, 1971). More generally, stream segregation appears to impair the perception of temporal sound relationships: once sounds fall in separate streams, it becomes very hard to perceive accurately their relative timing (for recent examples of this, see Roberts et al., 2008; Micheyl et al., 2010). Another line of evidence for a relationship between auditory streaming and perceptual performance comes from findings of elevated pitch-discrimination thresholds between consecutive ("target") tones when interfering tones are present, and are perceived as part of the same stream as the target sounds (e.g., Micheyl et al., 2005a).

8.3.1.2 Neural Bases for Streaming

Studies of the neural basis of auditory streaming in humans can be traced back to the EEG studies of Alain and Woods (1994) and Sussman et al. (1998, (1999). These authors used the mismatch negativity (MMN), a difference in event-related potentials (ERPs) between "deviant" and "standard" sound patterns in an "oddball" sound sequence containing random deviants among more frequent standards. Two conclusions emerged from these early studies. First, auditory streams are formed at or before the level at which the MMN is generated. Unfortunately, the neural generators of the MMN are difficult to localize precisely using scalp-surface recordings. They appear to involve a distributed cortical network, including contributions from temporal as well as frontal sources (e.g., Giard et al., 1990). This makes it difficult to ascertain, based on MMN data, whether auditory streaming is accomplished entirely in auditory cortex, or whether it requires neural structures beyond auditory cortex. To this day, a clear answer to this question is still lacking. Although some (fMRI) data suggest that brain areas beyond auditory cortex, such as the intraparietal sulcus, are differentially activated depending on whether the listeners perceives one stream or two (Cusack, 2005), it is not clear whether such differential activation participated in the formation of streaming percepts, or followed it.

The second conclusion that was reached in these early MMN studies is that focused auditory attention to the evoking sound sequence is not necessary for stream segregation. This conclusion was based on the observation that the MMN was present in participants who were reading a book during the presentation of the auditory stimuli, which suggested that the participants were not attentive to the auditory stimuli. This conclusion was challenged by Carlyon et al. (2001), who pointed out that the participants in those studies could devote some attention to the auditory stimuli, while they were reading. The same authors found that when listeners were given a difficult monaural auditory perception task while alternating tones were presented in the opposite ear, stream segregation apparently failed to build up, suggesting that focused auditory attention is required for stream segregation. An alternative interpretation of these psychophysical findings, however, is that switching attention back to the alternating-tone sequence caused a "resetting" of the perceptual state, from segregation back to integration. According to this interpretation, the results of Carlyon et al. (2001) are actually compatible with the view that sustained attention to the evoking sequence is not required for stream segregation. A more recent study by Sussman et al. (2007) specifically examined the effect of sustained attention and suggested that attention is not always required for stream segregation. Given that measuring perception in the absence of attention is virtually impossible, and that putative correlates of auditory streaming based on the MMN are not unquestionable, the debate concerning the role of attention in stream segregation will no doubt continue.

Although the MMN may provide an objective marker of auditory streaming in human auditory cortex, it provides limited insight into the neural basis for the phenomenon. Starting in 2001, a series of studies used multi- or single-unit recordings to investigate the neural mechanisms of auditory streaming at the level of primary cortex in macaque monkeys (*Macaca fascicularis*; Fishman et al., 2001; Micheyl et al., 2005a), bats (*Pteronotus parnellii*; Kanwal et al., 2003), and the European starlings (*Sturnus vulgaris*) field L2 (an avian analogue of mammalian primary auditory cortex; (Bee & Klump, 2004). Detailed reviews of these findings can be found in earlier publications (e.g., Micheyl et al., 2007; Snyder & Alain, 2007; Fay, 2008). The results of these studies have been be interpreted within the framework of a "tonotopic" model of auditory streaming. According to this model, a single stream is heard when the A and B tones excite the same or largely overlapping populations of frequency-selective neurons in A1. In contrast, when the A and B tones evoke spatially segregated patterns of activity in A1, two streams are heard. Importantly, the strength of neural responses to consecutive tones depends not just on frequency separation, but also on temporal parameters, with short intertone intervals promoting forward suppression (Fishman et al., 2001). The interaction between frequency-selectivity and forward-suppression effects appears, for the most part, to be consistent with the psychophysically observed dependence of auditory streaming on frequency separation and repetition rate. Moreover, multisecond adaptation of neural responses in primary cortex provides a relatively simple explanation for the buildup of stream segregation over time (Micheyl et al., 2005a).

It is important to note, however, that these conclusions were based on the results of experiments that involved strictly sequential (i.e., temporally nonoverlapping) tones. In a recent study, Elhilali et al. (2009) measured neural responses to sequences of alternating or synchronous tones in ferret A1. Their results indicate that, at least for relatively large frequency separations, the responses of A1 neurons tuned to one of the two frequencies present in the sequence do not differ appreciably depending on whether the tone at the other frequency is presented synchronously or sequentially. On the other hand, psychophysical experiments, which were performed in the same study, indicate that, in human listeners, the synchronous tones were usually heard as one stream, whereas the sequential tones evoked a strong percept of stream segregation. This indicates that while frequency selectivity, forward masking, neural adaptation, and contrasts in tonotopic population responses shape the neural representations of stimulus sequences in A1, there are aspects of the perception of auditory streams that are still not adequately explained by recordings from animal auditory cortex (Shamma & Micheyl, 2010).

One major caveat in the search for neural correlates of perceptual experience relates to the potentially confounding influence of stimulus differences (e.g., see Parker & Newsome, 1998). When different neural-response patterns are observed across different stimulus conditions, it is difficult to ascertain whether the differences reflect different percepts, or just different stimuli. This limits the conclusions of studies in which neural responses recorded under different stimulus conditions (such as different frequency separation, or intertone intervals) are compared to decide whether these neural differences reflect different percepts. One way to overcome this problem involves recording neural responses to physically constant but perceptually variable stimuli, and simultaneous measurements of the participant's percept. This approach has been used to identify neural correlates of conscious experience in the visual cortex in macaques during binocular rivalry (Logothetis & Schall, 1989).

A similar experimental strategy has been applied to study neural correlates of auditory streaming percepts at the cortical level in humans, using fMRI (Cusack, 2005; Kondo & Kashino, 2009), MEG (Gutschalk et al., 2005), or EEG (Snyder et al., 2006; Snyder et al., 2009). The results have led to the demonstration of changes in the BOLD signal or in MEG responses in auditory cortex (Gutschalk et al., 2005; Kondo & Kashino, 2009), in the thalamus (Kondo & Kashino, 2009), as well as in the intraparietal sulcus (Cusack, 2005), which appear to be related to changes in the listeners' percept (one stream vs. two streams). Further study is required to determine whether, and how, these changes in the level of neural (or metabolic) activity participate in the formation and perception of separate streams. It could be that some of the effects observed in these studies merely reflect attentional modulations triggered by different auditory percepts. An important goal for future studies of the neural basis of auditory streaming is to provide further clarity on this.

It was mentioned previously that auditory streaming can be observed with a wide variety of sounds. Relatively few studies have examined the neural basis of auditory streaming with stimuli other than pure tones. Deike et al. (2004) used fMRI to measure activity in human auditory cortex while listeners were presented with sequences of harmonic complex tones with alternating spectral envelopes, which were tailored to evoke organ-like and trumpet-like timbres. The results showed greater activation in the left but not in right auditory cortex during the presentation of sequences with alternating spectral envelopes and associated timbre, compared to the condition with a constant spectral envelope. The authors interpreted this result as evidence for a selective involvement of left auditory cortex during stream segregation based on timbre cues conveyed by spectral differences. Interestingly, Wilson et al. (2007) also observed greater activation in response to ABAB sequences of pure tones with a small or null A-B frequency separation than in response to sequences with a larger frequency separation, or slower sequences. They explained this in terms of forward suppression within frequency-specific neural populations. It is not entirely clear whether an equally simple explanation also holds for the findings of Deike et al.

To determine whether the effects observed in the fMRI study of Wilson et al. (2007) and in the MEG study of Gutschalk et al. (2005) were specific to pure tones, and critically dependent upon the responses of frequency-specific (tonotopic) mechanisms, Gutschalk et al. (2007) used both MEG and fMRI to measure auditory cortex activation by sequences of bandpass-filtered complex tones, which contained only unresolved harmonics in their passband. The results showed clear increases in auditory cortex activation as a function of the F0 difference between A and B tones in repeating AAAB sequences. Thus, F0 differences had an effect similar to that observed in previous studies based on frequency differences between pure tones, and this was the case even though tonotopic factors were eliminated, or at least, strongly reduced through the use of low F0s and bandpass-filtering. A relatively simple explanation for these observations could be based on neurons in auditory cortex that are sensitive to F0 or pitch (Bendor & Wang, 2005, 2006). The responses of these pitch-tuned neurons to rapidly presented tones may be as susceptible to forward suppression as those of frequency-tuned neurons in A1. From a more general perspective, forward suppression may enhance contrasts in the responses of sequentially activated neural populations tuned along other dimensions than just frequency or F0, such as amplitude-modulation rate. This provides a general neural mechanism for auditory streaming based on nontonotopic cues. A recent study extended these findings of neural correlates of auditory streaming based on nontonotopic cues to streaming based on interaural time differences (Schadwinkel & Gutschalk, 2010).

The EEG, MEG, and fMRI studies reviewed above raise interesting questions concerning the involvement of auditory and nonauditory cortical areas in auditory streaming. Intracranial studies and source analysis of human data (Liegeois-Chauvel et al., 1994; Gutschalk et al., 2004) indicate that the P1m and N1m are generated mostly in nonprimary auditory areas, including lateral HG, PT, and superior temporal gyrus (STG). Thus, the MEG data of Gutschalk et al. (2005) have been interpreted as evidence that neural responses in nonprimary cortex covary with listeners' percepts of auditory streaming. Gutschalk et al. found no evidence for modulation of neural responses outside of auditory cortex in their data. On the other hand, the fMRI data of Cusack (2005) showed no evidence for percept-dependent modulation of neural responses in auditory cortex. In that study, differential activation associated with the percept of one or two streams was only seen beyond auditory cortex, in the intraparietal sulcus (IPS). However, at the same time, Cusack found no significant change in auditory cortical activation dependent on the frequency separation between the A and B tones, whereas a later fMRI study by Wilson et al. (2007), which focused on auditory cortex, showed clear changes in fMRI activation in both HG and PT with increasing frequency separation. A possible explanation for the apparent discrepancy between these results stems from the consideration that the experimental design and analysis techniques used by Gutschalk et al. (2005) and Wilson et al. (2007) were better suited to capture stimulus- or percept-related auditory cortical activity than the whole-brain acquisitions used by Cusack (2005). Conversely, the approach used in the former studies was less well suited to the measurement of changes in neural activity outside of auditory cortex. Thus, these studies provide different windows on cortical activity during auditory streaming. Taken together, the results of these studies indicate that both auditory and nonauditory cortical areas are involved in auditory streaming. An important goal for future studies of the neural basis of auditory streaming will be to clarify the contributions of primary and secondary auditory cortex areas, and the possible role of areas outside auditory cortex (including the parietal and frontal lobes) in the generation of auditory streaming percepts. In addition, it will be important to determine more precisely the extent to which neural precursors of auditory streaming are already present (or not) in subcortical—and even, peripheral—levels of the auditory system.

8.3.2 Auditory Continuity

8.3.2.1 Psychophysics of Auditory Continuity

A pure tone that is interrupted for a few tens of milliseconds by silence is heard distinctly as discontinuous. However, if the silent gap is filled with noise, which overlaps spectrally with the tone and has a level sufficiently high to have masked

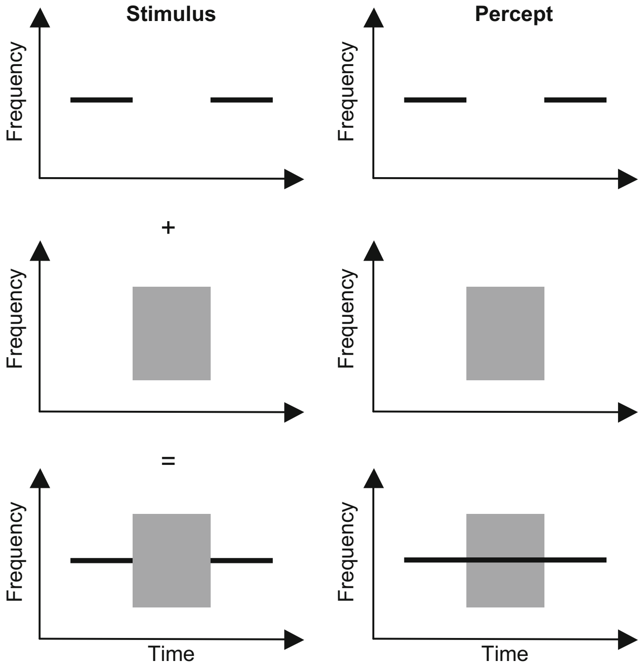


Fig. 8.4 Schematic illustration of stimuli and percepts in the auditory continuity illusion. (Top) A pure-tone that is physically interrupted for a few tens of milliseconds is perceived as clearly discontinuous. (Middle) A brief but temporally continuous noise band is heard as such. (Bottom) When the noise band is added to the pure-tone, in such a way that it fills the temporal gap in that tone, the tone is illusorily perceived as "continuing through" the noise, as if it were uninterrupted. For this auditory "continuity illusion" to occur, it is necessary that the noise be loud enough to mask the tone, if the tone was presented simultaneously with it

the tone (had it continued through the noise), listeners typically hear the tone as "continuing through" the noise, as if it were physically uninterrupted (Fig. 8.4) (Warren et al., 1972; Bregman & Dannenbring, 1977; Ciocca & Bregman, 1987). This phenomenon is known as the auditory "continuity illusion." The illusion bears some superficial resemblance to its visual homologue, in which two identically oriented line segments separated by a larger geometric figure (e.g., a rectangle) are perceived as a line behind the figure (Kanizsa & Gerbino, 1982). Illusory continuity may play an important role in everyday life, as sounds of interest are often interrupted by louder extraneous sounds in the environment (Warren et al., 1972); the illusion of continuity might help to counteract masking effects, and to maintain object coherence over time despite acoustic interference.

While parametric studies of the auditory continuity illusion have often employed pure tones (e.g., Riecke et al., 2009a), the illusion has also been demonstrated using other types of sounds than steady pure tones, including amplitude- or frequency-modulated tones or sweeps (Ciocca & Bregman, 1987; Kluender & Jenison, 1992;

Carlyon et al., 2004), and harmonic complex tones (Plack & White, 2000; Darwin, 2005). The occurrence of illusory continuity with speech sounds, which is usually referred to as "phonemic restoration" (Warren, 1970; Warren & Obusek, 1971; Carlyon et al., 2002), is of special interest. While phonemic restoration depends on acoustic characteristics (such as spectral similarity between the inducer and inducee) as well low-level sensory factors (such as peripheral or "energetic" masking) (Warren & Obusek, 1971), it is also strongly influenced by higher level factors specific to speech processing, such as phonemic expectations (Samuel, 1981).

Although auditory streaming and the continuity illusion have traditionally been studied separately, Tougas and Bregman (1990) pointed out that these two phenomena can be considered different aspects of a more general scene-analysis process, the function of which is to build accurate auditory representations. Therefore, the question of the relationship between the auditory continuity illusion and auditory streaming has been raised (Bregman & Dannenbring, 1973; Tougas & Bregman, 1990; Darwin, 2005). The empirical evidence so far appears to support the view that for illusory continuity to be perceived, the sound segments that precede and follow the interruption must be perceived as part of the same stream.

Finally, it is interesting to note that humans are not the only ones to experience the auditory continuity illusion. Behavioral studies have provided compelling evidence that primates also experience it (Miller et al., 2001; Petkov et al., 2003), serving as a basis for studies of neural correlates of the phenomenon in nonhuman species.

8.3.2.2 Neural Bases for Auditory Continuity

Two studies based on mammalian primary cortex demonstrated single neurons that responded similarly to continuous sounds, and to discontinuous sounds that were interrupted by intense noise—in such a way that they were heard as continuous. In the first study, Sugita (1997) measured the responses of cells in cat primary cortex to frequency glides. When these were interrupted by a silent gap, the responses were found to be considerably reduced, compared to uninterrupted glides. However, when the silent gap was filled with bandpass noise, to which the cells did not respond when this noise was presented in isolation, the response strength was restored. These results were interpreted as evidence that cells in primary cortex integrate information over time, providing a possible substrate for the percept of illusory continuity in primary cortex. A puzzling feature of this study, however, is that the noise was filtered into a remote frequency region, so that it did not overlap spectrally with the glide. Based on the results of several psychophysical studies in humans, one should not have expected the glide to be perceived as continuous under such stimulus conditions. The second study was performed by Petkov et al. (2007). These authors measured responses from single neurons to continuous and interrupted pure tones (with the interruption silent or filled with masking noise) in primary cortex of awake macaque monkeys. Under conditions in which the tone was either physically continuous, or heard as continuous (due to the insertion of noise in the gap), some neurons produced prominent onset responses at the onset of the tone, and either no response or a weak onset responses to the noise and second tone segments. In contrast, when the tone was interrupted by a silent gap, the second tone segment evoked a salient onset response in those same neurons. This pattern of results is consistent with the hypothesis that perceived illusory continuity is reflected in the responses of some neurons in primary cortex.

The first study devoted specifically to testing for neural correlates of the auditory continuity illusion in human auditory cortex was performed by Micheyl et al. (2003). These authors measured ERPs using an oddball paradigm involving four stimulus conditions, which were designed specifically to tease apart the relative contributions of stimulus-related and percept-related (illusory continuity) factors in the generation of the MMN. The pattern of results that was obtained in this study was interpreted as consistent with the hypothesis that, at the level at which the MMN is generated, perceived continuity is already reflected in the activity of neurons or neural populations. In addition, since the MMN is elicited even when participants are not actively attending to the stimuli (in this study, participants were watching a silent movie), these results suggest that the neural processes responsible for filling in gaps in sounds operate automatically, and do not require active attention to sound.

A limitation of the EEG study of Micheyl et al. (2003) is that the neural generators of the MMN could not be located precisely. This limitation was overcome in recent fMRI studies by Riecke et al. (2007, 2009b), which investigated neural correlates of the illusory continuity in human auditory cortex. In this study, participants in the scanner were presented with amplitude-modulated tones, which were either physically continuous or interrupted by a short temporal gap, which was either left empty of filled with a burst of noise. The fMRI data were analyzed, first, based on the physical characteristics of the stimuli, then, based on the ratings of perceived continuity provided by the listeners. These analyses showed differences in neural activation patterns evoked by physically identical stimuli in primary auditory cortex, depending on the listener's percept. These fMRI results are consistent with the above-described single-unit and EEG results in indicating the auditory continuity illusion is generated in or below the primary auditory cortex.

Finally, two recent studies have investigated neural correlates of the auditory continuity illusion produced by speech, or speech-like, stimuli in human listeners (Heinrich et al., 2008; Shahin et al., 2009). Unlike those reviewed in the preceding text, these studies did not focus exclusively on auditory cortex but used whole-brain fMRI. The first study (Heinrich et al., 2008) used synthetic vowels consisting of two formants (spectral peaks) occupying different spectral regions, which were presented either simultaneously (in which case the stimuli were perceived as speech-like sounds) or in alternation (in which case the stimuli were not recognized as speech sounds). The results revealed significant activation in posterior middle temporal gyrus (MTG) and superior temporal sulcus (STS)—two areas shown in previous studies to be involved specifically in speech processing—in the condition that involved simultaneous formants. Crucially, significantly greater activation was found in MTG in the condition involving alternating formants with high-level noise, which evoked the illusion, than in the condition involving alternating formants with high-level noise, which did not evoke the illusion. This outcome is consistent with the hypothesis that, at least for these stimuli, the auditory continuity illusion is generated prior to the level at which sound is processed specifically as speech. Interestingly, opposite activation patterns were observed in right and left HG containing primary cortex: more activation was observed in the condition in which the formants were alternating with silent gaps than in the condition in which the gaps were filled with noise, or the formants were continuous. It was suggested that neural activity in primary cortex depends primarily on stimulus onsets.

The second study (Shahin et al., 2009) used an elegant design to distinguish between two types of neural mechanisms of illusory continuity: (1) unconscious and largely bottom-up "repair" mechanisms, which contribute to restore missing sensory information automatically and (2) higher-level, error-detection-and-correction mechanisms, which compare bottom-up information with internally generated predictions, and are actually responsible for conscious percepts of illusory continuity. It was hypothesized that the latter mechanisms might recruit regions located outside of auditory cortex, such as left inferior frontal gyrus, which previous studies have indicated to play a role in the segmentation and recognition of acoustic sequences. The results of this study suggest that sensory-repair mechanisms take place in Broca's area, bilateral anterior insula, and the presupplementary motor area, whereas the mechanisms that are actually responsible for conscious percepts of illusory continuity for speech stimuli recruit the left angular gyrus (AG), STG, and the right STS. Overall, the results are consistent with the view that there exist two distinct paths in the brain, corresponding to two types of mechanisms that both contribute, more or less automatically, and more or less consciously, to the illusory continuity of speech stimuli.

To summarize, in the studies of single neurons above, EEG and fMRI responses concur to suggest that the auditory continuity illusion produced by pure tones interrupted by noise involves relatively automatic, early, and modality-specific neural mechanisms, which can be found in auditory cortex, and more specifically, primary cortex. On the other hand, studies using synthetic or natural speech indicate that brain regions located beyond auditory cortex are crucially involved in generating conscious percepts of illusory continuity for such stimuli, and raise questions as to whether neural activity in primary auditory cortex merely reflects physical stimulus properties, such as onsets.

8.3.3 Regularity and Deviance in Acoustic Sequences

Acoustic sequences contain information that increases as a function of the complexity of those sequences—more disordered sequences are less predictable so that each new sound in the sequence adds more information than a predictable sequence. The auditory system constantly needs to assess the statistical properties of streams of acoustic stimuli to understand the information contained and also to detect when one stream of information ends and another begins. Bayesian approaches to this problem treat the brain as an inference machine, which forms predictions from the statistical properties of sensory input and evaluates these predictions based on stored, experience-dependent templates or priors (Friston, 2003, 2005). Within this framework, the auditory system is constantly evaluating the incoming signal with respect to its statistical properties, from which it forms predictions that are the basis for detecting transitions in the auditory scene when the signal properties change.

The MMN paradigm has been enormously informative for auditory neuroscience; however, its simplistic nature limits inferences to the complex everyday acoustic environment. A few studies have attempted to address the processing of more complex statistical regularities and the expectancies that arise from them. For example, Bendixen et al. (2009) presented participants with isochronous tone sequences in which every other tone was a repetition of its predecessor while recording EEG activity. Thus, while the frequency of the first tone of such tone pairs was unpredictable, the second was predetermined and could be predicted. By infrequently omitting either the first or the second tone, the authors were able to test whether the auditory system either retrospectively fills in information (missing first tone) or prospectively predicts information (missing second tone). The results support the latter, showing that the omission of a predictable (but not the unpredictable) tone evoked a response that was similar to the response to the actual tone. This suggests that the auditory system preactivates the neural circuits for expected input. Overath et al. (2007) took this idea one step further: they used the entropy metric derived from information theory (Shannon, 1948) to create pitch sequences for which the statistical characteristics were held within probabilistic constraints, such that the general predictability of pitches within a pitch sequence was varied parametrically: for pitch sequences with high entropy, pitches were highly unpredictable and thus each pitch contributed new information, while pitch sequences with low entropy were more redundant because the general gist of pitches and pitch movement could be anticipated. From an information-theoretic perspective, high entropy pitch sequences require more computational demands than redundant pitch sequences, presumably because the auditory system tends to preactivate its neural circuitry (Bendixen et al., 2009), for example via "sparse" or "predictive" coding strategies (Friston, 2003, 2005). Using fMRI, the authors found that PT increased its activity as a function of stimulus entropy (or unpredictability). In contrast, a distributed frontoparietal network for retrieval of acoustic information operated independently of entropy. The results support the PT as an efficient neural engine or "computational hub" (Griffiths & Warren, 2002) that demands fewer computational resources to encode redundant signals than those with high information content.

Another approach to investigating how the auditory system encodes statistical regularities in sequences is to examine the effect of changing the stimulus statistics. Chait et al. (2008) used sequences of tone-pip stimuli that alternated in frequency either regularly or randomly and created transitions from the regular to the random pattern or vice versa. The former stimulus requires the discovery of a violation of regularity (regular to random), whereas the latter requires the detection of a new regularity (random to regular). Using MEG, the authors found that the temporal dynamics and morphology of the neural responses reveal distinct neural substrates in primary and nonprimary auditory cortex. This study and that of Overath et al. (2007)

demonstrate mechanisms in auditory cortex that differentiate between order and disorder (or predictability and randomness) in sound sequences.

The importance of stimulus statistics for auditory perception is not new. In fact, Shannon (1948) described his information theoretic approach of human communication with respect to letter probabilities in written language. In speech, transition probabilities between different phonemes adhere to constraints in a given language, and even very young infants are sensitive to these transition probabilities both in their native language (Saffran et al., 1996), an unfamiliar language (Pelucchi et al., 2009), as well as with more abstract pitch transition probabilities in short musical melodies (Saffran et al., 1999). Thus, the auditory system seems to group acoustic elements into meaningful entities, based on learned probabilities between those elements. This is a powerful strategy, as it allows both the grouping as well as segregation of auditory objects in a generic framework, as described earlier in this chapter.

8.4 Concluding Comments: Higher-Level Mechanisms

In the last section the idea was developed that auditory perception is generative: based on the communication between higher level areas that impose Bayesian priors on the incoming sensory information. The idea has been developed more fully in the visual system. For example, Rao and Ballard (1999) suggested an influential predictive coding model in which incoming sensory information in primary cortex is compared with a prediction signal from higher cortex and an error signal is communicated from primary cortex to higher cortex. Similar schemes have been suggested for the analysis of auditory objects too with prominent back projections from higher areas to lower areas: for example, auditory applications of reverse hierarchy theory discussed in Shamma (2008).

This chapter developed a "bottom up" approach based on simultaneous and then sequential grouping of object features that depends critically on processing in the primary and nonprimary auditory cortices, in addition to distinct areas beyond auditory cortex. Generative models represent efficient means of understanding the acoustic world and require effective connectivity between primary cortex and higher centers, and *systems* for the analysis of objects. Many of the experiments described have sought specific nodes in the system for the analysis of different aspects of objects, but a full understanding of the problem is likely to require systems identification (for a detailed discussion of this technique, see Griffiths et al., 2009b) to define these nodes and their effective connections during object analysis. The priors from higher areas might have a "label" like a word, or another type of stored meaning like a position in space. Generative models therefore immediately suggest ways in which the grouped patterns abstracted during object analysis can allow semantic processing.

Acknowledgments Griffiths is a Wellcome Trust Senior Clinical Fellow. Griffiths and Overath are supported by the Wellcome Trust, UK (WT091681MA). Micheyl is supported by the National Institute of Health, USA (R01 for Chris' NIH grant: DC07657).

References

Acoustical terminology. (1960). New York: American Standards Association.

Alain, C., & Woods, D. L. (1994). Signal clustering modulates auditory cortical activity in humans. *Perception and Psychophysics*, 56(5), 501–516.

Alain, C., & Izenberg, A. (2003). Effects of attentional load on auditory scene analysis. *Journal of Cognitive Neuroscience*, 15(7), 1063–1073.

Anstis, S., & Saida, S. (1985). Adaptation to auditory streaming of frequency-modulated tones. *Journal of Experimental Psychology: Human Perception and Performance*, 11(3), 257–271.

Bee, M. A., & Klump, G. M. (2004). Primitive auditory stream segregation: A neurophysiological study in the songbird forebrain. *Journal of Neurophysiology*, 92(2), 1088–1104.

Bee, M. A., & Micheyl, C. (2008). The cocktail party problem: what is it? How can it be solved? And why should animal behaviorists study it? *Journal of Comparative Psychology*, 122(3), 235–251.

Bendixen, A., Schröger, E., & Winkler, I. (2009). I heard that coming: event-related potential evidence for stimulus-driven prediction in the auditory system. *Journal of Neuroscience*, 29(26), 8447–8451.

Bendor, D., & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature*, 436(7054), 1161–1165.

Bendor, D., & Wang, X. (2006). Cortical representations of pitch in monkeys and humans. *Current Opinion in Neurobiology*, 16(4), 391–399.

Bizley, J. K., Walker, K. M., Silverman, B. W., King, A. J., & Schnupp, J. W. (2009). Interdependent encoding of pitch, timbre, and spatial location in auditory cortex. *Journal of Neuroscience*, 29(7), 2064–2075.

Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience*, 8(3), 389–395.

Bregman, A. S. (1990). Auditory scene analysis: The perceptual organisation of sound. Cambridge, MA: MIT Press.

Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, 89(2), 244–249.

Bregman, A. S., & Dannenbring, G. (1973). The effect of continuity on auditory stream segregation. *Perception and Psychophysics*, 13(2), 308–312.

Bregman, A. S., & Dannenbring, G. L. (1977). Auditory continuity and amplitude edges. *Canadian Journal of Psychology*, 31(3), 151–159.

Carlyon, R. P. (2004). How the brain separates sounds. *Trends in Cognitive Sciences*, 8(10), 465–471.

Carlyon, R. P., Cusack, R., Foxton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, 27(1), 115–127.

Carlyon, R. P., Deeks, J., Norris, D., & Butterfield, S. (2002). The continuity illusion and vowel identification. *Acta Acustica United with Acustica*, 88(3), 408–415.

Carlyon, R. P., Micheyl, C., Deeks, J. M., & Moore, B. C. (2004). Auditory processing of real and illusory changes in frequency modulation (FM) phase. *Journal of the Acoustical Society of America*, 116(6), 3629–3639.

Chait, M., Poeppel, D., Simon, J. Z. (2008). Auditory temporal edge detection in human auditory cortex. *Brain Research*, 1213, 78–90.

Chi, T., Ru, P., & Shamma, S. A. (2005). Multiresolution spectrotemporal analysis of complex sounds. *Journal of the Acoustical Society of America*, 118(2), 887–906.

Ciocca, V. (2008). The auditory organization of complex sounds. *Frontiers in Bioscience*, 13, 148–169.

Ciocca, V., & Bregman, A. S. (1987). Perceived continuity of gliding and steady-state tones through interrupting noise. *Perception and Psychophysics*, 42(5), 476–484.

Cusack, R. (2005). The intraparietal sulcus and perceptual organization. *Journal of Cognitive Neuroscience*, 17(4), 641–651.

Darwin, C. J. (2005). Simultaneous grouping and auditory continuity. *Perception and Psychophysics*, 67(8), 1384–1390.

Darwin, C. J., & Carlyon, R. P. (1995). Auditory Grouping. In B. C. J. Moore (Ed.), *Hearing* (pp. 387–424). San Diego: Academic Press.

de Cheveigné, A. (2005). Pitch perception models. In C. J. Plack, A. J. Oxenham, R. R. Fay & A. N. Popper (Eds.), *Pitch: Neural coding and perception* (pp. 169–233). New York: Springer-Verlag.

Deike, S., Gaschler-Markefski, B., Brechmann, A., & Scheich, H. (2004). Auditory stream segregation relying on timbre involves left auditory cortex. *NeuroReport*, 15(9), 1511–1514.

Dyson, B. J., & Alain, C. (2004). Representation of concurrent acoustic objects in primary auditory cortex. *Journal of the Acoustical Society of America*, 115(1), 280–288.

Elhilali, M., Ma, L., Micheyl, C., Oxenham, A. J., & Shamma, S. A. (2009). Temporal coherence in the perceptual organization and cortical representation of auditory scenes. *Neuron*, 61(2), 317–329.

Fay, R. R. (2008). Sound source perception and stream segregation in nonhuman vertebrate animals. In W. A. Yost, A. N. Popper & R. R. Fay (Eds.), *Auditory perception of sound sources* (pp. 307–323). New York: Springer.

Fishman, Y. I., & Steinschneider, M. (2010). Formation of auditory streams. In A. Rees & A. Palmer (Eds.), *The Oxford handbook of auditory science. The auditory brain*. Oxford: Oxford University Press.

Fishman, Y. I., Reser, D. H., Arezzo, J. C., & Steinschneider, M. (2001). Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hearing Research*, 151(1–2), 167–187.

Friston, K. (2003). Learning and inference in the brain. *Neural Networks*, 16(9), 1325–1352.

Friston, K. (2005). A theory of cortical responses. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 360(1456), 815–836.

Giard, M. H., Perrin, F., Pernier, J., & Bouchet, P. (1990). Brain generators implicated in the processing of auditory stimulus deviance: A topographic event-related potential study. *Psychophysiology*, 27(6), 627–640.

Griffiths, T. D., & Warren, J. D. (2002). The planum temporale as a computational hub. *Trends in Neurosciences*, 25(7), 348–253.

Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nature Reviews Neuroscience*, 5(11), 887–892.

Griffiths, T. D., Bamiou, D. E., & Warren, J. D. (2009a). Disorders of the auditory brain. In A. Rees & A. R. Palmer (Eds), *Oxford handbook of auditory science: The auditory brain* (pp. 509–542). Oxford: Oxford University Press.

Griffiths, T. D., Kumar, S., Von Kriegstein, K., Overath, T., Stephan, K. E., & Friston, K. J. (2009b). Auditory object analysis. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 367–381). Cambridge, MA: MIT Press.

Gutschalk, A., Patterson, R. D., Scherg, M., Uppenkamp, S., & Rupp, A. (2004). Temporal dynamics of pitch in human auditory cortex. *NeuroImage*, 22(2), 755–766.

Gutschalk, A., Micheyl, C., Melcher, J. R., Rupp, A., Scherg, M., & Oxenham, A. J. (2005). Neuromagnetic correlates of streaming in human auditory cortex. *Journal of Neuroscience*, 25(22), 5382–5388.

Gutschalk, A., Oxenham, A. J., Micheyl, C., Wilson, E. C., & Melcher, J. R. (2007). Human cortical activity during streaming without spectral cues suggests a general neural substrate for auditory stream segregation. *Journal of Neuroscience*, 27(48), 13074–13081.

Hall, D. A., & Plack, C. J. (2009). Pitch processing sites in the human auditory brain. *Cerebral Cortex*, 19(3), 576–585.

Heinrich, A., Carlyon, R. P., Davis, M. H., & Johnsrude, I. S. (2008). Illusory vowels resulting from perceptual continuity: A functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience*, 20(10), 1737–1752.

Kanizsa, G., & Gerbino, W. (1982). Amodal completion: Seeing or thinking? In J. Beck (Ed.), *Organization and representation in perception*. Hillsdale, NJ: Lawrence Erlbaum.

Kant, I. (1929). *A critique of pure reason*. Cambridge: Cambridge University Press.

Kanwal, J. S., Medvedev, A. V., & Micheyl, C. (2003). Neurodynamics for auditory stream segregation: Tracking sounds in the mustached bat's natural environment. *Network*, 14(3), 413–435.

Kluender, K. R., & Jenison, R. L. (1992). Effects of glide slope, noise intensity, and noise duration on the extrapolation of FM glides through noise. *Perception and Psychophysics*, 51(3), 231–238.

Kondo, H. M., & Kashino, M. (2009). Involvement of the thalamocortical loop in the spontaneous switching of percepts in auditory streaming. *Journal of Neuroscience*, 29(40), 12695–12701.

Krumbholz, K., Patterson, R. D., Seither-Preisler, A., Lammertmann, C., & Lutkenhoner, B. (2003). Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. *Cerebral Cortex*, 13(7), 765–772.

Kubovy, M., & Van Valkenburg, D. (2001). Auditory and visual objects. *Cognition.*, 80(1–2), 97–126.

Kumar, S., Stephan, K. E., Warren, J. D., Friston, K. J., & Griffiths, T. D. (2007). Hierarchical processing of auditory objects in humans. *PLoS Computational Biology*, 3(6), e100.

Liegeois-Chauvel, C., Musolino, A., Badier, J. M., Marquis, P., & Chauvel, P. (1994). Evoked potentials recorded from the auditory cortex in man: Evaluation and topography of the middle latency components. *Electroencephalography and Clinical Neurophysiology*, 92(3), 204–214.

Lipp, R., Kitterick, P., Summerfield, Q., Bailey, P. J., & Paul-Jordanov, I. (2010). Concurrent sound segregation based on inharmonicity and onset asynchrony. *Neuropsychologia*, 48(5), 1417–1425.

Logothetis, N. K., & Schall, J. D. (1989). Neuronal correlates of subjective visual perception. *Science*, 245(4919), 761–763.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, 58(3), 177–192.

Micheyl, C., Carlyon, R. P., Shtyrov, Y., Hauk, O., Dodson, T., & Pullvermuller, F. (2003). The neurophysiological basis of the auditory continuity illusion: A mismatch negativity study. *Journal of Cognitive Neuroscience*, 15(5), 747–758.

Micheyl, C., Carlyon, R. P., Cusack, R., & Moore, B. C. J. (2005a). Performance measures of auditory organization. In D. Pressnitzer, A. de Cheveigné, S. McAdams & L. Collet (Eds.), *Auditory signal processing: Physiology*, *psychoacoustics*, *and models* (pp. 203–211). New York: Springer.

Micheyl, C., Tian, B., Carlyon, R. P., & Rauschecker, J. P. (2005b). Perceptual organization of tone sequences in the auditory cortex of awake macaques. *Neuron*, 48(1), 139–148.

Micheyl, C., Carlyon, R. P., Gutschalk, A., Melcher, J. R., Oxenham, A. J., Rauschecker, J. P., Tian, B., & Courtenay Wilson, E. (2007). The role of auditory cortex in the formation of auditory streams. *Hearing Research*, 229(1–2), 116–131.

Micheyl, C., Hunter, C., & Oxenham, A. J. (2010). Auditory stream segregation and the perception of across-frequency synchrony. *Journal of Experimental Psychology: Human Perception and Performance*, 36(4), 1029–1039.

Middlebrooks, J. C., Clock, A. E., Xu, L., & Green, D. M. (1994). A panoramic code for sound location by cortical neurons. *Science*, 264(5160), 842–844.

Miller, C. T., Dibble, E., & Hauser, M. D. (2001). Amodal completion of acoustic signals by a nonhuman primate. *Nature Neuroscience*, 4(8), 783–784.

Miller, L. M., & Recanzone, G. H. (2009). Populations of auditory cortical neurons can accurately encode acoustic space across stimulus intensity. *Proceedings of the National Academy of Sciences of the USA*, 106(14), 5931–5935.

Moore, B. C. J., & Gockel, H. (2002). Factors influencing sequential stream segregation. *Acta Acustica United with Acustica*, 88, 320–333.

Nelken, I. (2004). Processing of complex stimuli and natural scenes in the auditory cortex. *Current Opinion in Neurobiology*, 14(4), 474–480.

Overath, T., Kumar, S., von Kriegstein, K., & Griffiths, T. D. (2008). Encoding of spectral correlation over time in auditory cortex. *Journal of Neuroscience*, 28(49), 13268–13273.

Overath, T., Cusack, R., Kumar, S., von Kriegstein, K., Warren, J. D., Grube, M., et al. (2007). An information theoretic characterisation of auditory encoding. *PLoS Biol*, 5(11), e288.

Overath, T., Kumar, S., Stewart, L., von Kriegstein, K., Cusack, R., Rees, A., & Griffiths, T. D. (2010). Cortical mechanisms for the segregation and representation of acoustic textures. *Journal of Neuroscience*, 30(6), 2070–2076.

Parker, A. J., & Newsome, W. T. (1998). Sense and the single neuron: probing the physiology of perception. *Annual Review of Neuroscience*, 21, 227–277.

Patterson, R. D. (2000). Auditory images: How complex sounds are represented in the auditory system. *Journal of the Acoustical Society of Japan*, 21, 183–190.

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, 36(4), 767–776.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Development*, 80(3), 674–685.

Petkov, C. I., O'Connor, K. N., & Sutter, M. L. (2003). Illusory sound perception in macaque monkeys. *Journal of Neuroscience*, 23(27), 9155–9161.

Petkov, C. I., O'Connor, K. N., & Sutter, M. L. (2007). Encoding of illusory continuity in primary auditory cortex. *Neuron*, 54(1), 153–165.

Plack, C. J., & White, L. J. (2000). Perceived continuity and pitch perception. *Journal of the Acoustical Society of America*, 108(3 Pt 1), 1162–1169.

Pressnitzer, D., & Hupé, J. M. (2006). Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. *Current Biology*, 16(13), 1351–1357.

Pressnitzer, D., Meddis, R., Delahaye, R., & Winter, I. M. (2001). Physiological correlates of comodulation masking release in the mammalian ventral cochlear nucleus. *Journal of Neuroscience*, 21(16), 6377–6386.

Pressnitzer, D., Sayles, M., Micheyl, C., & Winter, I. M. (2008). Perceptual organization of sound begins in the auditory periphery. *Current Biology*, 18(15), 1124–1128.

Puschmann, S., Uppenkamp, S., Kollmeier, B., & Thiel, C. M. (2009). Dichotic pitch activates pitch processing centre in Heschl's gyrus. *NeuroImage*.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1), 79–87.

Riecke, L., van Opstal, A. J., Goebel, R., & Formisano, E. (2007). Hearing illusory sounds in noise: Sensory-perceptual transformations in primary auditory cortex. *Journal of Neuroscience*, 27(46), 12684–12689.

Riecke, L., Mendelsohn, D., Schreiner, C., & Formisano, E. (2009a). The continuity illusion adapts to the auditory scene. *Hearing Research*, 247(1), 71–77.

Riecke, L., Esposito, F., Bonte, M., & Formisano, E. (2009b). Hearing illusory sounds in noise: The timing of sensory-perceptual transformations in auditory cortex. *Neuron*, 64(4), 550–561.

Roberts, B., Glasberg, B. R., & Moore, B. C. (2008). Effects of the build-up and resetting of auditory stream segregation on temporal discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, 34(4), 992–1006.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926–1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70, 27–52.

Samuel, A. G. (1981). Phonemic restoration: insights from a new methodology. *Journal of Experimental Psychology: General*, 110(4), 474–494.

Schadwinkel, S., & Gutschalk, A. (2010). Activity associated with stream segregation in human auditory cortex is similar for spatial and pitch cues. *Cerebral Cortex*, in press.

Shahin, A. J., Bishop, C. W., & Miller, L. M. (2009). Neural mechanisms for illusory filling-in of degraded speech. *NeuroImage*, 44(3), 1133–1143.

Shamma, S. (2008). On the emergence and awareness of auditory objects. *PLoS Biology*, 6(6), e155.

Shamma, S. A., & Micheyl, C. (2010). Behind the scenes of auditory perception. *Current Opinion in Neurobiology*, 20(3), 361–366.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27, 379–423 and 623–656.

Snyder, J. S., & Alain, C. (2007). Toward a neurophysiological theory of auditory stream segregation. *Psychological Bulletin*, 133(5), 780–799.

Snyder, J. S., Alain, C., & Picton, T. W. (2006). Effects of attention on neuroelectric correlates of auditory stream segregation. *Journal of Cognitive Neuroscience*, 18(1), 1–13.

Snyder, J. S., Holder, W. T., Weintraub, D. M., Carter, O. L., & Alain, C. (2009). Effects of prior stimulus and prior perception on neural correlates of auditory stream segregation. *Psychophysiology*, 46(6), 1208–1215.

Sugita, Y. (1997). Neuronal correlates of auditory induction in the cat cortex. *NeuroReport*, 8(5), 1155–1159.

Sussman, E., Ritter, W., & Vaughan, H. G. J. (1998). Attention affects the organization of auditory input associated with the mismatch negativity system. *Brain Research*, 789(1), 130–138.

Sussman, E., Ritter, W., & Vaughan, H. G. J. (1999). An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology*, 36(1), 22–34.

Sussman, E. S., Horvath, J., Winkler, I., & Orr, M. (2007). The role of attention in the formation of auditory streams. *Perception and Psychophysics*, 69(1), 136–152.

Tougas, Y., & Bregman, A. S. (1990). Auditory streaming and the continuity illusion. *Perception and Psychophysics*, 47(2), 121–126.

van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences.*University of Technology, Eindhoven.

von Helmholtz, H. L. F. (1885). *On the sensations of tone*, 4th (English translation 1912) ed. London: Longmans.

Warren, J. D., Jennings, A. R., & Griffiths, T. D. (2005). Analysis of the spectral envelope of sounds by the human brain. *NeuroImage*, 24(4), 1052–1057.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, 167(917), 392–393.

Warren, R. M., & Obusek, C. J. (1971). Speech perception and phonemic restorations. *Perception and Psychophysics*, 9, 358–362.

Warren, R. M., Obusek, C. J., & Ackroff, J. M. (1972). Auditory induction: Perceptual synthesis of absent sounds. *Science*, 176(39), 1149–1151.

Wilson, E. C., Melcher, J. R., Micheyl, C., Gutschalk, A., & Oxenham, A. J. (2007). Cortical FMRI activation to sequences of tones alternating in frequency: Relationship to perceived rate and streaming. *Journal of Neurophysiology*, 97(3), 2230–2238.

Chapter 9
Speech Perception from a Neurophysiological Perspective

Anne-Lise Giraud and David Poeppel

9.1 Introduction: Terminology and Concepts

Of all the signals human auditory cortex has to process, the one with the most compelling relevance to the listener is arguably speech. Parsing and decoding speech—the conspecific signal affording the most rapid and most precise transmission of information—must be considered one of the principal challenges of the auditory system. This chapter concentrates on what speech perception entails and what the constituent operations might be, emphasizing a neurophysiological perspective.

Research on speech perception is profoundly interdisciplinary. The questions range from (1) characterizing the relevant properties of the acoustic signal (*acoustic phonetics*, *engineering*) to (2) identifying the various (neurophysiological, neurocomputational, psychological) subroutines that underlie the perceptual analysis of the signal (*neuroscience*, *computation*, *perceptual psychology*) to (3) understanding the nature of the representation that forms the basis for creating meaning (*linguistics*, cognitive *psychology*). The entire process comprises—at least—a mapping from mechanical vibrations in the ear to abstract representations in the brain.

One terminological note merits emphasis. *Speech perception* refers to the mapping from sounds to internal linguistic representations (roughly, words). This is not coextensive with *language comprehension*. Language comprehension can be mediated by ear (speech perception), but also by eye (reading, sign language, lip reading), or by touch (Braille). Thus, *speech perception proper comprises a set of auditory processing operations prior to language comprehension*. The failure to distinguish

A.-L. Giraud (✉)
Inserm U960, Département d'Etudes Cognitives, Ecole Normale Supérieure,
29 rue d'Ulm, 75005 Paris, France
e-mail: anne-lise.giraud@ens.fr

D. Poeppel
Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, USA
e-mail: david.poeppel@nyu.edu

D. Poeppel et al. (eds.), *The Human Auditory Cortex*, Springer Handbook of Auditory Research 43, DOI 10.1007/978-1-4614-2314-0_9,

between speech and language has led to much unfortunate confusion; because the goal is to identify the critical component operations that underlie speech (and ultimately language) comprehension, a meticulous subdivision of the relevant cognitive science and linguistics terminology is essential. How does this translate into research practice? Insofar as we are interested in studying properties of words that are central to comprehension, but abstract and independent of the input modality, we would aim to find features that are stable across auditory, visual, or tactile presentation. In contrast, when we study speech perception, we are interested in the attributes that underlie the transformation from an acoustic signal to the possible internal representations. Because speech perception can thus be viewed as a subroutine of language comprehension in which the computation of meaning is not required, it can be approached, at least in part, by investigating the perception of isolated speech sounds (e.g., vowels or consonant-vowel syllables) or single words.

Current models of speech perception (and the associated neurobiological literature) tend to derive from studies of the perception of single speech sounds, syllables, or words. For example, the phenomenon of categorical perception (Liberman et al., 1967) as well as the work on vowel inventories (e.g., Näätänen et al., 1997) has stimulated an enormous literature on understanding sublexical perceptual processes. Aspects related to categorical perception have been examined and reviewed in detail (e.g., Harnad, 1987) and continue to motivate neurobiological studies on category formation and processing (Sharma & Dorman, 1999; Blumstein et al., 2005, Chang et al., 2010). Similarly, the experimental research on spoken word recognition (e.g., using tasks such as lexical decision, gating, priming, or shadowing) has laid the basis for prominent perception models, including the cohort model (Gaskell & Marslen-Wilson, 2002), the lexical access from spectra approach (Klatt, 1989), the TRACE model (McClelland & Elman, 1986), and others.

The literature has been ably reviewed and examined from different perspectives (Hawkins 1999; Cleary & Pisoni, 2001; Pardo & Remez, 2006), including from a slightly more linguistically motivated vantage point (Poeppel & Monahan, 2008; Poeppel et al., 2008). In addition, the related body of engineering research on automatic speech recognition has added important insights; this work, too, has been extensively reviewed (Rabiner & Juang, 1993). A recent book-length treatment of speech perception bridging acoustics, phonetics, neuroscience, and engineering is provided in Greenberg and Ainsworth (2006).

The goal of this chapter is to focus explicitly on the processing of naturalistic, connected speech, that is, *sentence level speech analysis*. The motivation for focusing on connected speech is threefold. First, there is a renewed interest in focusing on ecologically relevant, naturalistic stimulation. The majority of laboratory research places participants in artificial listening situations with peculiar task demands (e.g., categorical perception, lexical decision, etc.), typically unrelated to what the listener does in real life. That the execution of such task demands has a modulatory influence on the outcome of neurobiological experiments and leads to serious interpretive problems has been discussed at length (e.g., Hickok & Poeppel, 2000, 2004, 2007). Second, investigating speech perception using sentence level stimuli has a prominent history worth linking to; however, only in the last decade is it playing an increasing role in cognitive neuroscience and neurophysiology (e.g., Scott et al., 2000; Luo & Poeppel, 2007; Friederici et al., 2010). Early and formative contributions to understanding speech research were made by focusing on signal-to-noise ratio and intelligibility of sentences. An influential monograph by Miller (1951) summarized some of this early work, which is also deeply influenced by engineering demands (for a recent discussion, see Allen, 2005). This early work highlighted the relevance of temporal parameters for speech. Third, some of the most provocative new insights into speech processing come from data on listeners exposed to sentence level input. As mentioned, the focus on single speech sounds placed a large emphasis on the relevance of detailed spectral cues (e.g., formant patterns) and short-term temporal cues (e.g., formant transitions) on recognition performance. In contrast, the recent work on sentence-level stimuli (i.e., materials with a duration exceeding 1–2 s), and using experimental task demands such as intelligibility, demonstrate the fundamental importance of long-term temporal parameters of the acoustic signal. A growing literature in human auditory neuroscience has identified attributes of the system that underlie processing of communicative signals at this level. Important new principles have been discovered.

The chapter proceeds as follows. First, some of the essential features of speech are outlined. Next, the properties of auditory cortex that reflect its sensitivity to these features are reviewed (Section 9.2) and current ideas about the processing of connected speech are discussed (Section 9.3). The chapter closes with a summary of speech processing models at a larger scale that attempt to capture many of these phenomena in an integrated manner (Section 9.4).

9.2 Processing Speech as an Acoustic Signal

9.2.1 Some Critical Cues

Naturalistic, connected speech is an aperiodic but quasi-rhythmic acoustic signal with complex spectrotemporal modulations, that is, complex variations of the frequency pattern over time. Figure 9.1 illustrates two useful ways to visualize the signal: as a waveform (A) and as a spectrogram (B). The waveform represents energy variation over time—the input that the ear actually receives. The outlined “envelope” (thick line) reflects that there is a temporal regularity in the signal at relatively low modulation frequencies. These modulations of signal energy (in reality, spread out across a filterbank) are below 20 Hz and peak roughly at a rate of 46 Hz (Steeneken & Houtgast, 1980; Elliott & Theunissen, 2009). From the perspective of what auditory cortex receives as input, namely the modulations at the output of each frequency channel of the filterbank that constitutes the auditory periphery (cf. Hall and Barker, Chapter 7), these energy fluctuations can be characterized by the modulation spectrum (Kanedera et al., 1999; Greenberg & Kingsbury, 1997). Importantly, these slow-energy modulations correspond roughly to the syllabic structure (or syllabic “chunking”) of speech. The syllabic structure as reflected by the envelope, in turn,

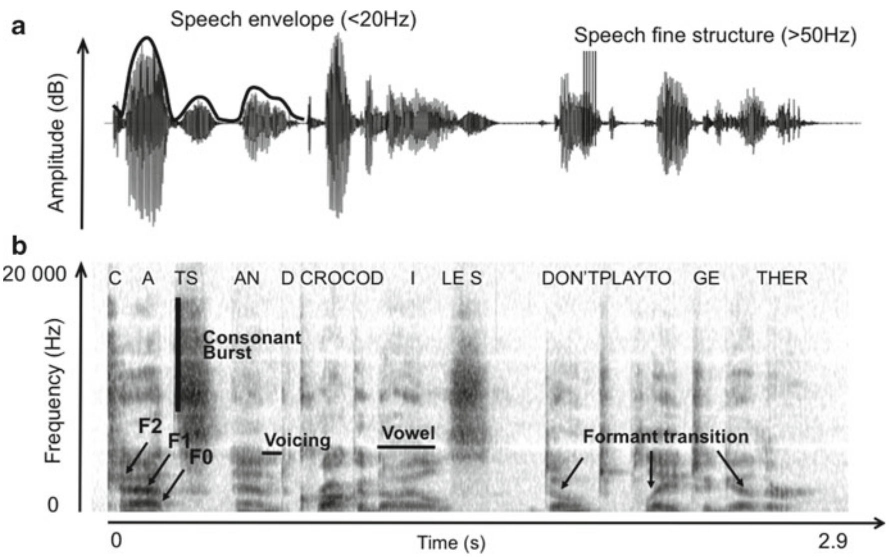


Fig. 9.1 Waveform (**a**) and spectrogram (**b**) of the same sentence uttered by a male speaker. Some of the key acoustic cues in speech comprehension are highlighted in black

is perceptually critical because it signals the speaking rate, it carries stress and tonal contrasts, and cross-linguistically the syllable can be viewed as the carrier of the linguistic (question, statement, etc.) or affective (happy, sad, etc.) prosody of an utterance (Rosen, 1992). As a consequence, a high sensitivity to envelope structure and envelope dynamics is critical for successful speech perception.

The second analytic representation, the spectrogram, decomposes the acoustic signal in the frequency, time, and amplitude domains (Fig. 9.1B). Textbook summaries often suggest that the human auditory system captures frequency information between 20 Hz and 20 kHz (and such a spectrogram is plotted here), but most of the information that is extracted for effective recognition lies below 8 kHz. It is worth remembering that speech transmitted over telephone landlines contains a much narrower bandwidth (200–3600 Hz) and is comfortably understood by normal listeners. A number of critical acoustic features can be identified in the spectrogram. The faintly visible vertical stripes represent the glottal pulse, which reflects the speaker's fundamental frequency, F0. This can range from approximately 100 Hz (male adult) to 300 Hz (child). The horizontal bands of energy show where in frequency space a particular speech sound is carried. The spectral structure thus reflects the articulator configuration. These bands of energy include the formants (F1, F2, etc.), definitional of vowel identity; high-frequency bursts associated, for example, with frication in certain consonants (e.g., /s/, /f/); and formant transitions that signal the change from a consonant to a vowel or vice versa.

The fundamental frequency (F0) conveys important cues about the speaker, for example, gender and size, and its modulation signals the prosodic contour of an utterance (including, sometimes, lexical boundaries) and intonation (stress); F0 can also convey phonetic information (in tonal languages). The formants, mainly F1, F2, and F3, define the identity of vowels. The ratio between F1 and F2 is relatively characteristic of each vowel. Cues for vowel discrimination are thus mainly of spectral nature, if we assume that the auditory system computes F1/F2 ratios. It has also been suggested that the ratio of F3/F2 and F3/F1 can be computed online; this measure has high utility for speaker normalization (Monahan & Idsardi, 2010). It goes without saying that to compute such ratios, the auditory system must first extract the frequency structure of the sound.

Consonants are often associated with more transient acoustic properties, and with a broader spectral content. The energy bursts underlying consonants can range from partial obstructions of air flow (e.g., in fricatives such as /f/) to the release of energy after full occlusion (e.g., in stop consonants /p/, /t/, or /k/). Consonants can be discriminated either by the spectral content of their initial burst, that is, by the fast formant transitions that bridge consonant and vowels, or by the presence of voicing (Rosen, 1992), which corresponds to vocal chord vibrations occurring before and during the consonant burst. This means consonants are (or can be) discriminated on the basis of a mixture of spectral and temporal cues. All of these cues are present within the acoustic fine structure, that is, signal modulations at faster rates, say above 50 Hz. The capacity of the auditory brain to capture the speech fine structure is therefore important to recovering important details of the signal.

There exist excellent summaries of acoustic phonetics. Some emphasize the aspect of the productive apparatus (Stevens, 1998); others highlight a cross-linguistic perspective (Laver, 1994). There is a large body of data on the acoustic correlates of different attributes of speech, covered in dedicated textbooks (e.g., Pickett, 1999). Based on this brief and selective summary, two concepts merit emphasis: first, the extended speech signal contains critical information that is modulated at rates of less than 20 Hz, with the modulation peaking around 5 Hz. This low-frequency information correlates closely with the syllabic structure of connected speech. Second, the speech signal contains critical information at modulation rates higher than, say, 50 Hz. This rapidly changing information is associated with fine spectral changes that signal speech sound identity and other relevant speech attributes. Thus, there exist *two surprisingly different timescales concurrently at play in the speech signal*. This important issue is taken up in the text that follows.

Notwithstanding the importance of the spectral fine structure, there is a big caveat: speech can be understood, in the sense of being intelligible in psychophysical experiments, when the spectral content is replaced by noise and only the envelope is preserved. Importantly, this manipulation is done in separate bands across the spectrum, for example, as few as four separate bands (e.g., Shannon et al., 1995). Speech that contains only envelope but no fine structure information is called vocoded speech (Faulkner et al., 2000). Compelling demonstrations that exemplify this type of signal decomposition (Shannon et al., 1995; Ahissar et al., 1995; Smith et al., 2001) illustrate that the speech signal can undergo radical alterations and distortions and yet remain intelligible.

Such findings have led to the idea that the temporal envelope, that is, temporal modulations of speech at relatively slow rates, is sufficient to yield speech comprehension (Scott et al., 2006; Loebach & Wickesberg, 2008; Souza & Rosen, 2009). When using stimuli in which the fine structure is compromised or not available at all, envelope modulations below 16 Hz appear to suffice for adequate intelligibility. The remarkable comprehension level reached by most patients with cochlear implants, in whom about 15–20 electrodes replace 3000 hair cells, remains the best empirical demonstration that the spectral content of speech can be degraded with tolerable alteration of speech perception (Roberts et al., 2011). A related demonstration showing the resilience of speech comprehension in the face of radical signal impoverishment is provided by sine-wave speech (Remez et al., 1981). In these stimuli both envelope and spectral content are degraded but enough information is preserved to permit intelligibility. Typically sine-wave speech preserves the modulations of the three first formants, which are themselves replaced by sine-waves centered on F0, F1, and F2. In sum, *dramatically impoverished stimuli remain intelligible insofar as enough information in the spectrum is available to convey temporal modulations at appropriate rates*.

9.2.2 Sensitivity of Auditory Cortex to Speech Features

9.2.2.1 Sensitivity to Frequency

This section reviews the equipment of auditory cortex to process spectral and temporal cues relevant to speech. Primary auditory cortex (A1) is organized as a series of adjacent territories (cf. Clarke and Morosan, Chapter 2), which retain cochlear tonotopy, much like visual cortex is organized as series of retinotopic regions (cf. Hall and Barker, Chapter 7). This means that the spectral content of speech signals that is physically decomposed by the basilar membrane in the cochlea and encoded in primary auditory neurons (cochlear filters) is still place-coded at the level of core auditory cortex, and possibly in some adjacent territories. A place code can be important to discriminate speech sounds that differ with respect to their spectral content. Tonotopic maps are organized in auditory cortex as multiple "mirrors," resulting in an alternation of regions coding high and low frequencies (Formisano et al., 2003; Petkov et al., 2006). One of these functionally early auditory territories seems to be specifically involved in the processing of periodicity pitch (Patterson et al., 2002; Bendor & Wang 2006; Nelken et al., 2008), which corresponds to a sensation of tonal height conveyed by the temporal regularity of a sound, rather than by its audiofrequency content (see Griffiths et al., 2010). This region is located in the most lateral part of Heschl's gyrus overlapping with a region that is sensitive to very low frequency sounds, that is, the frequencies that correspond to pitch percepts, usually referred to as "the pitch domain" (for some discussion, see Hall and Barker, Chapter 7). Experiments using magnetoencephalography (MEG) have implicated the same area when pitch is constructed binaurally (Chait et al., 2006), extending the role of such an area to pitch analysis more broadly. The reason for the clustering of periodicity pitch and other pitch responses within this region is not well understood. A possible and parsimonious explanation could be that auditory neurons (not only cortical) with very low characteristic frequencies (CFs) respond equally well to an input from a cochlear filter with very low CF, and to the modulation at CF rate of other cochlear filters. In cortex, such an overlap can be envisaged as a transition between place and temporal coding principles (cf. Cariani and Micheyl, Chapter 13). Accordingly, the pitch domain corresponds to the lowest edge of the range of frequencies that can be decomposed by the basilar membrane's physical properties. With respect to speech processing, the pitch center should play an essential role in coding speaker identity and prosody/intonation contour. In line with this, functional magnetic resonance imaging (fMRI) studies in humans show, on the one hand, that the pitch center is more developed in right than left auditory cortex (Zatorre & Gandour, 2008), and on the other hand that identity of both vowels and speakers is better represented in right temporal cortex (Formisano et al., 2008), even though strong interactions across cortical hemisphere are necessary to complete complex speaker recognition tasks (von Kriegstein et al., 2010).

9.2.2.2 Sensitivity to Time

Most neurons in primary auditory cortex are sensitive to temporal properties of acoustic stimuli. Their discharge pattern easily phase-locks to pulsed stimuli of up to about 40–60 Hz (Bendor & Wang, 2007; Middlebrooks, 2008; Brugge et al., 2009). Yet, this ability is limited compared to subcortical neurons that can phase-lock to much higher rates. The ability to represent the temporal modulation of sounds by an "isomorphic" response pattern that precisely mimics the stimulus temporal structure with the discharges (Bendor & Wang, 2007) decreases from the periphery to auditory cortex. Whereas thalamocortical fibers can phase-lock up to around 100 Hz, neurons in the inferior colliculus, superior olive, and cochlear nucleus are able to follow even faster acoustic rates (Giraud et al., 2000; Joris et al., 2004). Thus, there is a dramatic temporal down-sampling from subcortical to cortical regions— and what follows from this architectural feature of cellular physiology is the need for different neural coding strategies. For acoustic modulations faster than 30–40 Hz, auditory cortical neurons respond only at the onset of stimulus, with remarkable precision (Abeles, 1982; Heil, 1997a, b; Phillips et al., 2002). In awake marmosets, Wang and colleagues identified two main categories of auditory cortical neurons. Whereas "synchronized" (phase-locking) neurons use a faithful temporal code (isomorphic) to represent stimulus temporal modulation, "unsynchronized" neurons use a rate code. In each of these categories, Bendor and Wang (2006) describe neurons that respond either by increasing (positive monotonic) or decreasing (negative monotonic) their discharge rate with stimulus modulation. Synchronized neurons that are able to phase-lock to the stimulus are essentially found in primary auditory cortex (A1). When moving away from A1, the proportion of unsynchronized "onset" neurons increases. Their response in several dimensions, that is, the amount of spikes per time unit, the delay between stimulus and response onset, the duration of the spike train (Bendor & Wang, 2006), and the precise spike-timing (Kayser et al., 2010), may be used to form abstract temporal information and to perform more elaborate and integrated computations, such as speech segmentation, grouping, etc. (Wang, 2007).

While phase-locking seems to saturate around 40 Hz in A1, Elhilali et al. (2004) observed that primary auditory neurons can follow stimulus modulations at faster rates (up to 200 Hz) when fast modulations ride on top of a slow modulations. With respect to speech, this ability means that when carried by the speech envelope, aspects of the fine structure can be "isomorphically" encoded by auditory cortical neurons. This suggests that there may be two different mechanisms for encoding slow and fast temporal modulations (Ding & Simon, 2009). Slow-amplitude modulations gate fast phase-locking properties, because slow modulations permit a periodic reset of synaptic activity and a regeneration of the pool of neurotransmitters (synaptic depression hypothesis). Although periodic synaptic regeneration is plausible, one could question why individual auditory neurons would have fundamentally different properties and biophysical limitations than subcortical auditory neurons. It is conceivable that the specificity of auditory cortical neurons lies in the fact that they are more massively embedded in large corticocortical networks, which requires that they not only faithfully follow and code input but also temporally structure output transmission. In sum, *the role of the auditory cortex is not only to efficiently represent the auditory input efficiently*, *but also*, *and perhaps primarily*, *to convert input structure into a code that will possibly be matched with other types of representations*. As exposed in the text that follows, ensemble *neuronal oscillations may help by temporally structuring neuronal output and facilitating the "packaging" and transformations to more abstract neural codes* and representations, and pooling together neuronal ensembles according to endogenous principles.

9.2.2.3 Sensitivity to Spectrotemporal Modulations

Speech signals are characterized by modulations in both spectral and temporal domains. Two separate possible codes to represent complex stimuli such as speech have been implicated in the preceding text, a place code for spectral modulations and a temporal code for temporal modulations. Whether spectral and temporal modulations are encoded by a single or by distinct mechanisms remains an open question. The idea of a single code for spectrotemporal modulations is supported by the presence of neurons that respond to frequency modulations but not amplitude modulations (Gaese & Ostwald, 1995) and by complex responses to spectrotemporal modulations (Schönwiesner & Zatorre, 2009; Pienkowski & Eggermont, 2010). Luo et al. (2006, 2007b) and Ding and Simon (2009) tested, based on MEG recordings in human listeners, whether FM and AM used the same coding principles. Figure 9.2 schematizes the stimulus configuration and the hypothesized neural coding strategies (see legend). The authors argue that if coding equivalence (or similarity) is the case, cortical

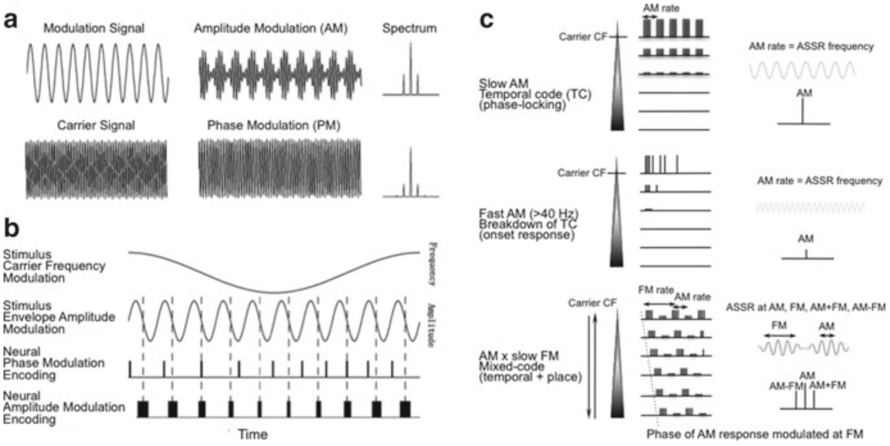


Fig. 9.2 Principles of amplitude and frequency modulations encoding in auditory cortex (Luo et al., Journal of Neurophysiology [2006], used with permission of APS). (**a**) In radio engineering modulation is used to encode acoustic stimuli, which can be either amplitude (AM; upper row) or phase modulated (PM; second row). (**b**) Proposals for neural AM and PM encoding. A stimulus is made of a frequency varying signal (upper row) and an amplitude modulation (second row). Using a PM encoding (third row), a neuron fires one spike per stimulus envelope cycle (dotted line) and the firing precise timing (phase) depends on the carrier frequency. Alternatively, using AM encoding (last row), a neuron changes its firing rate according to the instantaneous frequency of the carrier, while keeping constant the firing phase. (**c**) AM coding is illustrated in more detail in three different conditions, slow AM (upper row), fast AM (second row), and when AM and FM covary (last row). CF, characteristic frequency

responses as assessed by MEG should be the same when the carrier of slow AM is rapidly frequency modulated, or when a slowly changing carrier sound is amplitude modulated at fast rate (AM–FM comodulation experiments). Yet, they observed that only the phase of fast AM auditory responses (auditory steady state responses at 40 Hz) is modulated by slow FM, while both the phase and the amplitude of fast FM auditory responses (auditory steady state responses at 40 Hz) are modulated by slow AM. That AM and FM interact nonlinearly is beyond doubt. However, the mere fact that the spectral place-coding present in several auditory territories plays a more important role in FM processing than in AM processing could account for the asymmetry in the results. Whereas FM, by hypothesis, is encoded by a combination of place and temporal coding, AM is mostly encoded by temporal coding. Figure 9.2 depicts a model to characterize how AM and FM, critical features of speech signals, may plausibly be encoded, based on processing units that have a tonotopic axis and incorporate distinct thresholds for temporal stimulus modulations.

The asymmetric response pattern to fast and slow AM/FM might also depend on coding differences for fast and slow modulations. Whereas very slow frequency modulations are perceived as pitch variations, fast modulations are perceived as varying loudness. On the other hand, slow-amplitude modulations are perceived as variations of loudness, whereas fast modulations are perceived as roughness, or flutter, or pitch. These sharp perceptual transitions could be underpinned by both the size and the place of the population recruited by each of these stimulus types. Whereas slow FM presumably allows for both a temporal and spatial segregation of cortical responses, entailing distinct percepts varying in pitch, fast FM presumably phase-locks together at FM the entire population stimulated by the varying carrier. A slight jitter in phase-locking could then account for the roughness of the sensation. In a similar way, fast AM is possibly no longer perceived as variations of loudness when the ability of neurons to phase-lock is overridden (beyond 40 Hz). Flutter (and then pitch sensations) for AM higher than 40 Hz superimposed on the primary spectral content of the modulated sound might reflect the additional excitation of neurons with very low CF (pitch neurons).

The spectral place-code, the transition from phase-locking to rate-coding for higher stimulus rates, and ensemble neuronal behavior, that is, the size of the population targeted by a stimulus, provide enough representational complexity to account for nonlinear neuronal responses to spectrotemporal acoustic modulations without invoking a specific AM/FM code.

9.2.2.4 Sparse Representations in the Auditory Cortex

The described response properties in auditory cortex need to be interpreted with caution. Electrophysiological recordings necessarily rely on a selection of neurons, a selection that is often biased toward units that *fire* in response to auditory stimuli (Atencio et al., 2009). Many neurons, however, are silent. The picture that arises from electrophysiological studies is one of "dense" coding because we extrapolate population behavior from a few recorded neurons. Hromádka and Zador (2009) argue that no more than 5% of auditory neurons fire above 20 spikes/s at any instant. These authors suggest that, rather than "dense," auditory responses are "sparse" and highly selective, which permits more accurate representations and a better discrimination of auditory stimuli. Sparse coding implies the existence of population codes relying strongly on topographical organization and spatial patterns. The behavioral relevance of such *mesoscopic* cortical organization has been recently demonstrated using functional neuroimaging (Formisano et al., 2008; Eger et al., 2009). Rather than looking at the mean of the response to repeated stimulation, these methods analyze the variance across trials and show that what differs from one trial to the next is meaningfully represented in the spatial pattern of the response. This spatial pattern can be distributed across functional areas, for example, across tonotopic auditory areas (Formisano et al., 2008; Chang et al., 2010). The bottom line of these studies is that percepts are individually encoded in mesoscopic neural response patterns on a millimeter–centimeter scale. The notion of hierarchical processing across several tonotopically organized functional regions is somewhat deemphasized by this new perspective, which is more compatible with an analysis-by-synthesis or Bayesian view in which higher and lower processing stages conspire to generate a percept (elaborated in the last section).

9.3 Cortical Processing of Speech as a Continuous Stream

Experimental research on the neural basis of speech has tended to focus on processing individually presented speech sounds, such as vowels, syllables, or single words. This approach has led to good progress, and the findings underpin most current models of speech. That being said, a large part of naturalistic speech comes at the listener as a continuous stream, in phrases and sentences, and not well “prepackaged” in perceptual units of analysis. Indeed, this *segmentation* problem remains a major challenge to contemporary models of adult and child speech perception as well as automatic speech recognition. Interestingly, in psychophysical research on speech, especially through the 1950s, a large body of work studied speech perception and intelligibility using phrasal or sentential stimuli (see, e.g., Miller [1951] for a summary of many experiments and Allen [2005] for a review of the influential work of Fletcher and others). There exist fascinating findings based on that work, for example, on the role of signal-to-noise ratio, but the feature that arises is that speech as a continuous signal has principled and useful temporal properties that merit attention and that may play a key role in the problem of speech parsing and decoding.

9.3.1 The Discretization Problem

In natural connected speech, speech cues are embedded in a continuous acoustic flow. Their on-line analysis and representation by spatial, temporal and rate neural encodings (see Cariani and Micheyl, Chapter 13) needs to be read out (decoded) by mechanisms that are unlikely to be continuous. The first step of these neural parsing and read-out mechanisms should be the discretization of the continuous input signal and its initial neural encoding. The generalization that perception is “discrete” has been motivated and discussed in numerous contexts (e.g., Pöppel, 1988; Van Rullen & Koch, 2003). There is an important distinction between *temporal integration* versus *discretization*, which for expository purposes is glossed over in this chapter.

One particular hypothesis about a potential mechanism for chunking speech and other sounds is discussed here, namely that cortical oscillations could be efficient instruments of auditory cortex output discretization, or discrete sampling. Neural oscillations reflect synchronous activity of neuronal assemblies that are either intrinsically coupled or coupled by a common input. They are typically measured in animal electrophysiology by local field potential recordings (review: Wang, 2010). The requirements for measuring oscillations and spiking activity are different. When spiking is looked for, the experimenter typically tracks a response to a stimulus characterized by a fast and abrupt increase in firing rate. Oscillations, on the other hand, can be observed in the absence of stimulation, and are modulated by stimulation in a less conspicuously causal way. The selection bias is therefore much stronger when measuring spikes than oscillations, because spiking reflects activity of either a single neuron or a small cluster of neurons selective to certain types of

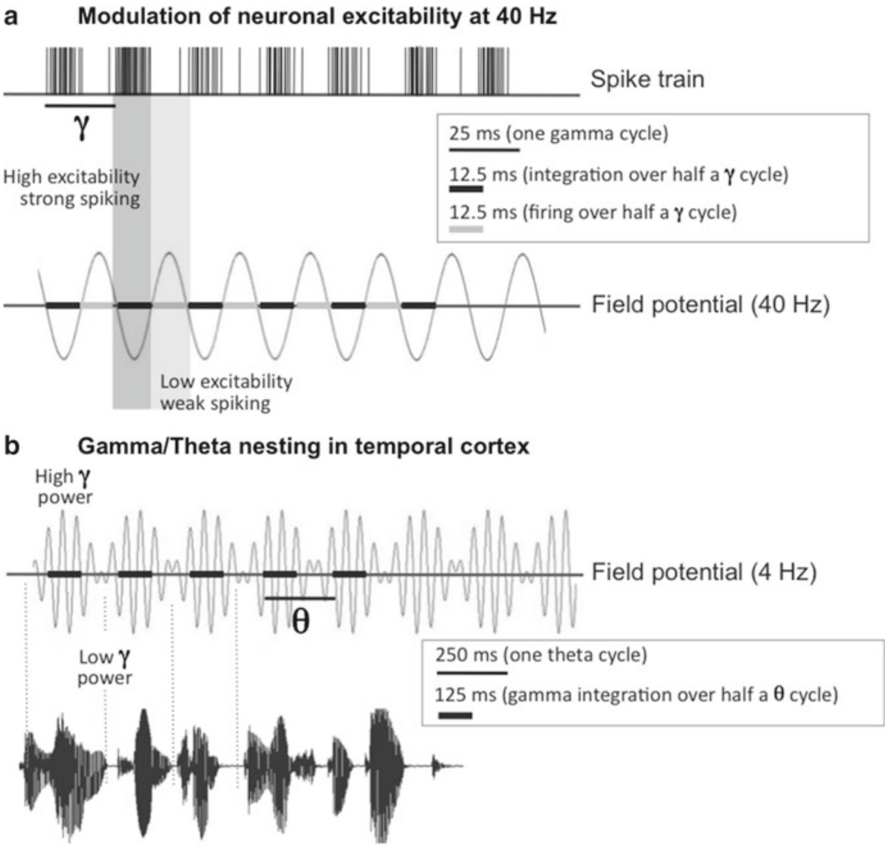

a
Modulation of neuronal excitability at 40 Hz
Spike train
γ
High excitability
strong spiking
25 ms (one gamma cycle)
12.5 ms (integration over half a γ cycle)
12.5 ms (firing over half a γ cycle)
Field potential (40 Hz)
Low excitability
weak spiking
b
Gamma/Theta nesting in temporal cortex
High γ
power
Field potential (4 Hz)
θ
Low γ
power
250 ms (one theta cycle)
125 ms (gamma integration over half a θ cycle)

Fig. 9.3 The temporal relationship between speech and brain oscillations. (**a**) Gamma oscillations periodically modulate neuronal excitability and spiking. The hypothesized mechanism is that neurons fire for about 12.5 ms and integrate for the rest of the 25-ms time window. Note that these values are approximate, as we consider the relevant gamma range for speech to lie between 28 and 40 Hz. (**b**) Gamma power is modulated by the phase of theta rhythm (about 4 Hz). Theta rhythm is reset by speech resulting in keeping the alignment between brain rhythms and speech bursts

stimuli and ready to fire at the right moment. Cortical oscillations are proposed to shape spike-timing dynamics and to impose phases of high and low neuronal excitability (Britvina & Eggermont, 2007; Schroeder and Lakatos, 2009a, b; Kayser et al., 2010). The assumption that it is oscillations that cause spiking to be temporally clustered derives from the observation that spiking tends to occur in the troughs of oscillatory activity (Womelsdorf et al., 2007). The principle is illustrated in Figure 9.3A. It is also assumed that spiking and oscillations do not reflect the same aspect of information processing. Whereas spiking reflects axonal activity, oscillations are said to reflect mostly dendritic synaptic activity (Wang, 2010).

Neuronal oscillations are ubiquitous in cerebral cortex and other brain regions, for example, hippocampus, but they vary in strength and frequency depending on their location and the exact nature of their neuronal generators (Mantini et al., 2007).

In human auditory cortex, at rest, approximately 40 Hz activity (low gamma band) is strong and can be measured using stereotactic electroencephalography (EEG) in epileptic patients, MEG, or concurrent EEG and fMRI (Morillon et al., 2010). Neural oscillations in this range are endogenous in the sense that one can observe a spontaneous grouping of spikes at approximately 40 Hz even in the absence of acoustic stimulation. This gamma activity is thought to be generated by a ping-pong interaction between pyramidal cells and inhibitory interneurons (Borgers et al., 2005; Borgers & Kopell, 2008), or even just among interneurons that are located in superficial cortical layers (Tiesinga & Sejnowski, 2009). In the presence of a stimulus, this patterning at gamma frequencies becomes more pronounced, and clustered spiking activity is propagated to higher hierarchical processing stages (Arnal et al., 2011). Input to auditory cortex is conveyed by thalamocortical fibers contacting cells in layer IV. Unlike visual cortex, auditory cortical layer IV does not contain spiny stellate cells, which are the primary target of thalamocortical input, but rather pyramidal cells (Binzegger et al., 2007; da Costa & Martin, 2010). Whereas spiny stellates are small neurons with a modest dendritic tree, forming a horizontal coat of interdigitated ramifications, pyramidal cells are essentially vertical elements, reaching far below and above the layer where their cell bodies are found. Although it is unclear why cortical canonical microcircuits might be differently organized in the auditory and visual cortices (see Atencio et al., 2009), it is possible that this more vertical architecture emphasizes sequential/hierarchical processing over spatial integrative processing, meeting more closely critical requirements of speech processing, where analysis of the temporal structure is as important as spectral analysis.

By analogy with the proposal of Elhilali et al.(2004) that fast responses are gated by slower ones, it is interesting to envisage this periodic modulation of spiking by ensemble oscillatory activity as an endogenous mechanism to ensure sustained excitability of the system. This endogenous periodicity, however, could also reflect the alternation of dendritic integration and axonal transmission, which needs to be slowed down in the cortex due to the large amount of data to integrate, and the relatively long time lags between inputs signaling a common single event, possibly even through different sensory channels. In ecological situations, speech perception relies on the integration of visual and auditory inputs that are naturally shifted by about 100 ms (see van Wassenhove and Schroeder, Chapter 11). Integration of audiovisual speech requires data accumulation over a larger time window than the one allowed for by gamma oscillations. Such integration could occur under the patterning of oscillations in the theta range. In the next section, a potential role of theta activity in speech processing is thus outlined.

9.3.2 Speech Analysis at Multiple Timescales

Based on linguistic, psychophysical, and physiological data as well as conceptual considerations, it has been proposed that speech is analyzed in parallel at multiple timescales (Poeppel, 2001, 2003; Boemio et al., 2005; Poeppel et al., 2008). The central idea is that both local-to-global and global-to-local types of analyses are carried out concurrently (multitime-resolution processing). The concept is related to reverse hierarchy theories of perception (Hochstein & Ahissar, 2002; Nahum et al., 2008). The principal motivations for such a hypothesis are twofold. First, a single, short temporal integration window that forms the basis for hierarchical processing, that is, increasingly larger temporal analysis units as one ascends the processing system, fails to account for the spectral and temporal sensitivity of the speech processing system and is hard to reconcile with behavioral performance. Second, the computational strategy of analyzing information on multiple scales is widely used in engineering and biological systems, and the neuronal infrastructure exists to support multiscale computation (Canolty & Knight, 2010). According to the view summarized here, speech is chunked into segments of roughly featural or phonemic length, and then integrated into larger units, as segments, diphones, syllables, words. In parallel, there is a fast global analysis that yields coarse inferences about speech (akin to Stevens' 2002 "landmarks" hypothesis), and that subsequently refines segmental analysis. Segmental and suprasegmental analyses could be carried out concurrently and "packaged" for parsing and decoding due to neuronal oscillations at different rates. Considering a mean phoneme length of about 25–80 ms and a mean syllabic length of about 150–300 ms, dual-scale segmentation is assumed to involve two sampling mechanisms, one at about 40 Hz (or, more broadly, in the low gamma range) and one at about 4 Hz (or in the theta range). Electrophysiological evidences in favor of this hypothesis are discussed later.

Schroeder and Lakatos (2009a, b) argue that oscillations determine phases of high and low excitability on pyramidal cells. This means that with a period of approximately 25 ms, gamma oscillations provide a 10- to 15-ms window for integrating spectrotemporal information (low spiking rate) followed by a 10- to 15-ms window for propagating the output (high spiking rate) (see, for illustration Fig. 9.3A.). However, a 10- to15-ms window of integration might be too short to characterize an approximately 50 ms phoneme. This raises the question of how many gamma cycles are required to encode phonemes correctly. This question has so far only been addressed using computational modeling (Shamir et al., 2009). Using a pyramidal interneuron gamma (PING) model of gamma oscillations (Borgers et al., 2005) that modulate activity in a coding neuronal population, Shamir et al. (2009) show that the shape of a sawtooth input signal designed to have the typical duration and amplitude modulation of a diphone (~50 ms; typically a consonant–vowel or vowel–consonant transition) can correctly be represented by three gamma cycles, which act as a three-bit code. This code has the required capacity to distinguish different shapes of the stimulus and is therefore a plausible means to distinguish between phonemes. That 50-ms diphones could be correctly discriminated with three gamma cycles suggests that phonemes could be sampled with one/two gamma cycles. This issue is critical, as *the frequency of neural oscillations in the auditory cortex might constitute a strong biophysical determinant with respect to the size of the minimal acoustic unit that can be manipulated for linguistic purposes*.

The notion of speech analysis at multiple timescales is useful because it allows the move from strictly hierarchical models of speech perception (e.g., Giraud & Price, 2001) to more complex models in which simultaneous extraction of different acoustic cues permits simultaneous high-order processing of different information from the same input signal. That speech *should* be analyzed in parallel at different timescales derives, among other reasons, from the observation that articulatory–phonetic phenomena occur at different timescales. It was noted previously (Fig. 9.1) that the speech signal contains events of different durations: short energy bursts and formant transitions occur within a 20- to 80-ms timescale, whereas syllabically carried information occurs over 150–300 ms. The processing of both types of events could be accounted for either by a hierarchical model in which smaller acoustic units (segments) are concatenated into larger units (syllables) or by a parallel model in which both temporal units are extracted independently, and then combined. A degree of independence in the processing of long (slow modulation) and short (fast modulation) units is observed at the behavioral level. For instance, speech can be understood well when it is first segmented into units up to 60 ms and when these local units are temporally reversed (Saberi & Perrott, 1999; Greenberg & Arai, 2001). This observation rules out the idea that speech processing relies solely on hierarchical processing of short and then larger units, as the correct extraction of short units is not a prerequisite for comprehension. Overall, there appears to be a grouping of psychophysical phenomena such that some cluster at thresholds of approximately 50 ms and below and others cluster at approximately 200 ms and above (a similar clustering is observed for temporal properties in vision; Holcombe 2009). Importantly, nonspeech signals are subject to similar thresholds. For example, 15–20 ms is the minimal stimulus duration required for correctly identifying upward versus downward FM sweeps (Luo et al., 2007a). By comparison, 200-ms stimulus duration underlies loudness judgments. In sum, physiological events at related scales form the basis for processing at that level. Gamma oscillations, for example, could act as an integrator such that all events occurring within about 15 ms are grouped, whereas events occurring within the next 15 ms are suppressed. Although it may sound inefficient to suppress half of the acoustic structure, an oscillatory mechanism could reflect a tradeoff between accurate signal extraction/representation and its on-line transmission to levels higher in the hierarchy, as well as ensuring the sustained excitability of the system.

9.3.3 Alignment of Neuronal Excitability with Meaningful Speech Events

An important requirement of the computational model mentioned previously (Shamir et al., 2009) is that ongoing gamma oscillations are phase-reset, for example, by a population of onset excitatory neurons. Without this onset signal the performance of the model drops. Ongoing intrinsic oscillations appear to be effective as a segmenting tool only if they *align* with the stimulus. Schroeder and colleagues suggest that gamma and theta rhythms work together, and that the phase of theta oscillations determines the power and possibly also the phase of gamma oscillations (see Fig. 9.3B; Schroeder & Lakatos, 2008). This relationship is referred to as "nesting." Electrophysiology suggests that theta oscillations can be phase-reset by several means, in particular through multimodal corticocortical pathways (Arnal et al., 2009), but most probably by the stimulus onset itself. The largest cortical auditory evoked response measured with EEG and MEG, about 100ms after stimulus onset, could correspond to the phase reset of theta activity (Arnal et al., 2011). This phase reset would align the speech signal and the cortical theta rhythm, the proposed instrument of speech segmentation into syllable/word units. As speech is strongly amplitude modulated at the theta rate, this would result in aligning neuronal excitability with those parts of the speech signals that are most informative in terms of energy and spectrotemporal content (Fig. 9.3B; Giraud and Poeppel, submitted). There remain critical computational issues, such as the means to get strong gamma activity at the moment of theta reset. Recent psychophysical research emphasizes the importance of aligning the acoustic speech signal with the brain's oscillatory/quasi-rhythmic activity. Ghitza and Greenberg (2009) demonstrated that comprehension can be restored by inserting periods of silence in a speech signal that was made unintelligible by time-compressing it by a factor of 3. The mere fact of adding silent periods to speech to restore an optimal temporal rate, which is equivalent to restoring "syllabicity," improves performance even though the speech segments that remained available are not more intelligible. Optimal performance is obtained when 80-ms silent periods alternate with 40-ms time-compressed speech. These time constants allowed the authors to propose a phenomenological model involving three nested rhythms in the theta (5 Hz), beta, or low gamma (20–40 Hz) and gamma (80 Hz) domains (for extended discussion, see Ghitza, 2011).

9.3.4 Multitime-Resolution Processing: Asymmetric Sampling in Time

Poeppel (2003) attempted to integrate and reconcile several of the strands of evidence: first, speech signals contain information on at least two critical timescales, correlating with segmental and syllabic information; second, many nonspeech auditory psychophysical phenomena fall in two groups, with integration constants of approximately 25–50 ms and 200–300 ms; third, both patient and imaging data reveal cortical asymmetries such that both sides participate in auditory analysis but are optimized for different types of processing in left versus right; and fourth, crucially for the present chapter, neuronal oscillations might relate in a principled way to temporal integration constants of different sizes. Poeppel (2003) proposed that there exist hemispherically asymmetric distributions of neuronal ensembles with preferred shorter versus longer integration constants; these cell groups "sample" the input with different sampling integration constants (Fig. 9.4A). Specifically, left

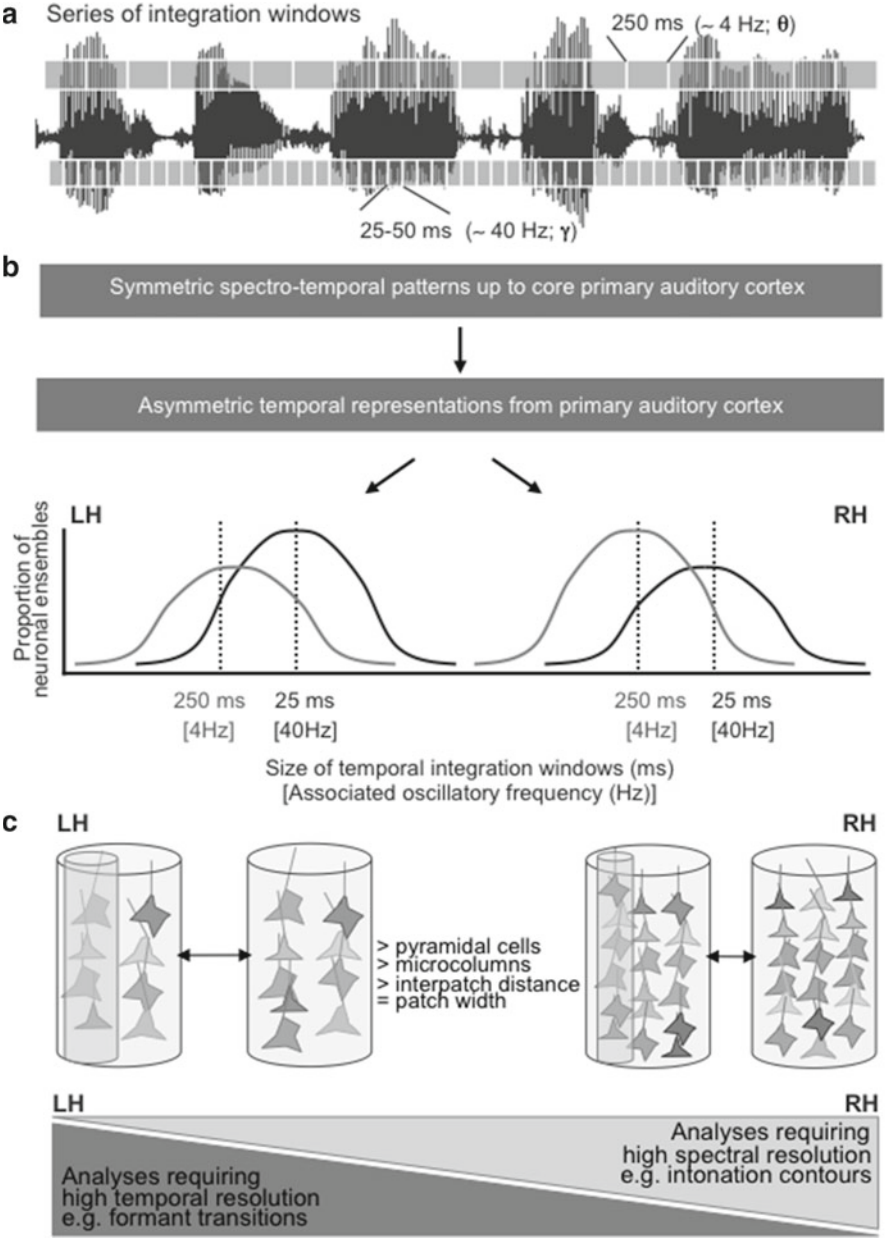

a
Series of integration windows
250 ms (~ 4 Hz; θ)
25-50 ms (~ 40 Hz; γ)
b
Symmetric spectro-temporal patterns up to core primary auditory cortex
Asymmetric temporal representations from primary auditory cortex
LH
RH
Proportion of neuronal ensembles
250 ms [4Hz]
25 ms [40Hz]
250 ms [4Hz]
25 ms [40Hz]
Size of temporal integration windows (ms) [Associated oscillatory frequency (Hz)]
c
LH
RH
> pyramidal cells
> microcolumns
> interpatch distance
= patch width
LH
RH
Analyses requiring high spectral resolution e.g. intonation contours
Analyses requiring high temporal resolution e.g. formant transitions

Fig. 9.4 (**a**) Temporal relationship between the speech waveform and the two proposed integration timescales (in ms) and associated brain rhythms (in Hz). (**b**) Proposed mechanisms for asymmetric speech parsing: left auditory cortex (LH) contains a larger proportion of neurons able to oscillate at gamma frequency than the right one (RH). (**c**) Differences in cytoarchitectonic organization between the right and left auditory cortices. Left auditory cortex contains larger pyramidal cells in superficial cortical layers and exhibits bigger microcolumns and a larger patch width and interpatch distance auditory cortex has a relatively higher proportion of short term (gamma) integrating cell groups, whereas right auditory cortex has a larger proportion of long term (theta) integrating neurons (Fig. 9.4B). As a consequence, left hemisphere auditory cortex is better equipped for parsing speech at the segmental scale, and right auditory cortex for parsing speech at the syllabic timescale. This hypothesis, referred to as the asymmetric sampling in time (AST) theory, is illustrated in Figure 9.4 and accounts for a variety of psychophysical and functional neuroimaging results that show that left temporal cortex responds better to many aspects of rapidly modulated speech content while right temporal cortex responds better to slowly modulated signals including music, voices, and other sounds (Zatorre et al., 2002; Warrier et al., 2009). A difference in the size of the basic integration window between left and right auditory cortices would explain speech functional asymmetry by a better sensitivity of left auditory cortex to information carried in fast temporal modulations that convey, for example, phonetic cues. A specialization of right auditory cortex to slower modulations would grant it a better sensitivity to slower and stationary cues such as harmonicity and periodicity (Rosen, 1992) that are important to identify vowels, syllables, and thereby speaker identity. The AST theory is very close, in kind, to the spectrotemporal asymmetry hypothesis promoted by Zatorre (e.g., Zatorre et al., 2002; Zatorre & Gandour, 2008).

As mentioned above, the underlying physiological hypothesis is that left auditory cortex contains a higher proportion of neurons capable of producing gamma oscillations than right auditory cortex. Conversely, right auditory cortex contains more neurons producing theta oscillations. Consistent with this proposal, Hutsler and Galuske (2003) showed that the microcolumnar organization is different in the left and right auditory cortices (Fig. 9.4C). Left auditory cortex contains larger pyramidal cells in layer III and larger microcolumns. It could be the case that larger pyramidal cells produce oscillations at higher rates because the larger the cell the stronger the membrane conductance and the faster the depolarization/repolarization cycle. Pyramidal cell conductance may play a role in setting the rhythm at which excitatory/inhibitory circuits (PING) oscillate. This hypothesis, however, has to be verified using computational models.

To evaluate the plausibility of this model, four types of data are required. First, temporal integration over the short timescale (for both speech and nonspeech auditory signals) must be demonstrated. Second, evidence of temporal integration over the longer time scale is necessary. There exists a body of such evidence, some of which is reviewed by Poeppel (2003). Pitch judgments versus loudness judgments exemplify the two timescales, as do segmental versus syllabic processing timescales. Third, the information on these two timescales should interact, to yield perceptual objects that reflect the integrated properties of both modulation rates. This has not been widely tested, but there is compelling behavioral evidence in favor, discussed briefly later. Finally, there should be cerebral asymmetries in the cortical response properties, which are summarized.

Relevant psychophysical data testing interactions across timescales are sparse, but several studies have attempted to understand the relative contributions of different modulation rates. Elliott and Theunissen (2009) provide data showing that there are interactions across bands with restricted temporal modulation frequencies, although they did not explicitly test the ranges of interest here. Chait et al. (submitted) show a striking interaction of two selected bands of speech signals in dichotic speech conditions: when both low (<8 Hz) and high (25–40 Hz) signals are presented concurrently, listeners' performance exceeds the predicted linear combination values, suggesting a clear interaction between the timescales of interest. Further, Saoud et al. (submitted) observed that speech comprehension is both faster and more accurate when the low-rate temporal envelope (0–4 Hz) of bisyllable words is presented through the left ear and the high temporal envelope (28–40 Hz) is presented to the right ear relative to the reverse dichotic situation. These results suggest (1) that the two timescales carry information that interacts synergistically to yield higher intelligibility representations of the input signal and (2) that comprehension is better when each auditory cortex receives speech information in a temporal format that matches its intrinsic oscillatory capacity. Recent fMRI evidence supports this conclusion (Saoud et al., 2012).

Despite a limited understanding of the psychophysics, a large number of imaging and neurophysiological studies have addressed the cerebral asymmetry predictions. For example, consistent with AST, Boemio et al. (2005), using temporally extended stimuli built from short segments of different durations, showed a striking rightwards asymmetry in superior temporal sulcus (STS) for these nonspeech stimuli when longer time segments were used (e.g., 300 ms), compared to the short-time-structure signals (e.g., 25 ms). Similarly, Overath et al. (2008) showed a significant rightward lateralization for auditory stimuli with increasing length of spectrotemporal time windows. Zaehle et al. (2004) tested speech and nonspeech signals and observed robust leftward lateralization for rapidly modulated auditory signals. Jamison et al. (2006) used nonspeech signals in an fMRI design and observed the predicted left/rapid–right/slow associations. The predictions have been tested for speech and nonspeech, and pitting spectral against temporal processing advantages (e.g., Obleser et al., 2008), including even in newborns (Telkemeyer et al., 2009). By and large, the predicted associations hold up well, and there is emerging consensus that temporal parameters of the sort discussed here play a central role in decoding auditory signals in the cortex.

Are the predicted asymmetric sampling properties truly architectural features of the system, or are the observed asymmetries driven into the system by properties of the stimuli employed? To verify that the sound analysis asymmetries are systemic properties, Giraud and colleagues (2007) measured the distribution of neuronal oscillations in subjects not exposed to input, that is, in a passive resting state. Using combined EEG/fMRI at rest, they discovered a stronger expression of gamma rhythm in left auditory cortex and a stronger expression in theta rhythm in right auditory cortex (Fig. 9.5A). Control analyses included analyses of other frequency bands in the alpha and low and high beta range. For these frequency bands there were no significant EEG/fMRI correlations in auditory cortex at rest and no detectable asymmetry.

The left hemisphere dominance of gamma activity at rest was confirmed using a detailed anatomical approach in another concurrent EEG/fMRI data set (Morillon

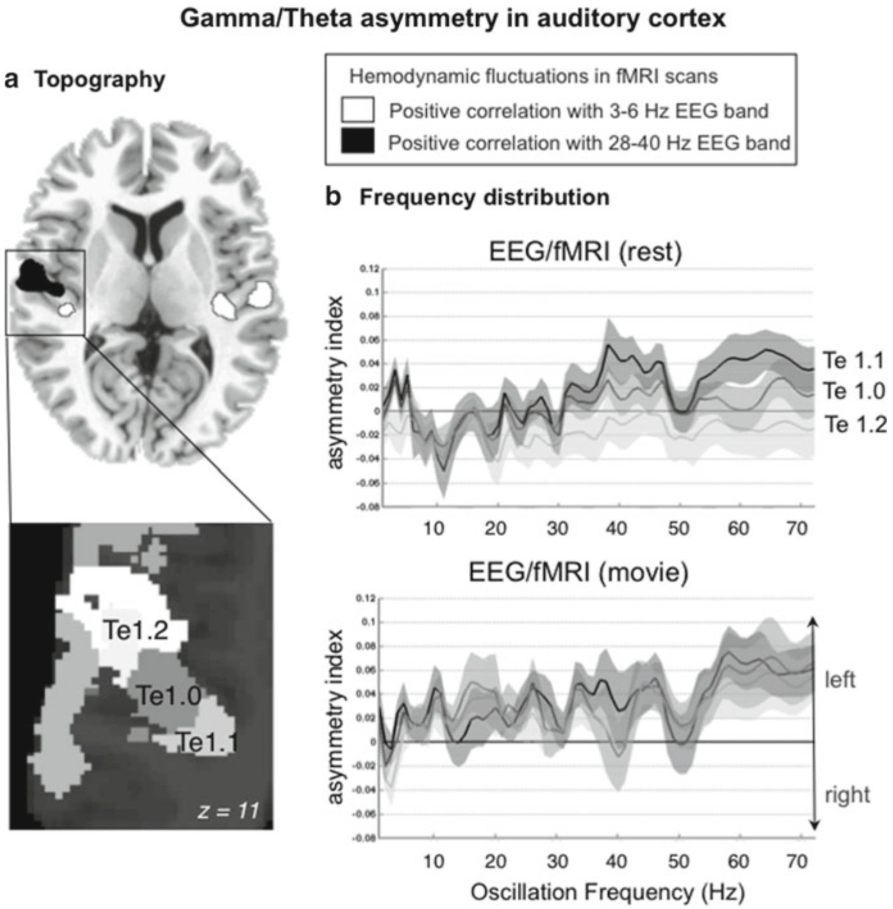


Fig. 9.5 (**a**) Experimental evidence for an asymmetry in cortical oscillations in the left and right auditory cortices at rest using combined EEG/fMRI (after Giraud et al., 2007). (**a**) Topographical distribution of EEG/fMRI coupling in the theta and low gamma bands. Note that both rhythms are expressed on both sides, but that a right/left dissociation can be seen at appropriate statistical threshold. (**b**) Correlations between EEG power and fMRI bold signal in three different cytoarchitectonic territories of Heschl's gyrus at rest and when subjects were watching a spoken movie (after Morillon et al., 2010). Asymmetry in the strength of EEG/fMRI correlation was maximal in Te 1.1 at rest (mostly within the gamma range) and increased in all three territories during audio-visual stimulation

et al., 2010). There, fMRI time series were extracted from various cytoarchitectonic territories along Heschl's gyrus and correlated with power variations of EEG over its entire spectrum (1–72 Hz). These data showed that the left dominance in spontaneous expression of gamma activity arises from the most posteromedial part of Heschl's gyrus (Te 1.1), and that it declines along its posteromedial to anterolateral axis (Fig. 9.5B). Because EEG/fMRI correlations are rather weak, these data were compared to MEG data at rest, from sensors that were pretested to be most responsive to auditory input. The latter analyses confirmed the left-dominance of gamma rhythm at rest. However, unlike previous results, both the region of interest based on the EEG/fMRI approach and the MEG data did not give a consistent picture of spontaneous theta activity. The variance across experimental data underscores that more experiments are needed to validate, invalidate, or augment the AST proposal.

The EEG/fMRI experimental data show that oscillations in the delta band (1–3 Hz) become right-dominant during linguistic processing, while most other rhythms including beta activity become strongly left-dominant (Fig. 9.5B, lower panel). The delta/low theta rhythm has the temporal properties to underlie prosodic processing, as it corresponds to integration of speech signals in approximately 500-ms windows. This rate would be ideal to mediate prosodic operations such as extracting intonation contours indicative of speaker's emotional states, illocutionary intent, etc. It is thus possible that rather than theta, it is the delta rhythm that is predominantly right lateralized, while gamma and theta rhythms jointly underlie speech parsing in left auditory cortex. There is a lot of work in progress regarding this unresolved question.

9.4 Large-Scale Neurocognitive Models of Speech Processing

9.4.1 Emerging Consensus: Functional Neuroanatomic Models

Although the perceptual analysis of speech is rooted in the different anatomic subdivisions of auditory cortex in the temporal lobe, speech processing involves a large network that includes areas in parietal and frontal cortices, the relative activations of which strongly depend on the task performed. Several reviews have synthesized the state-of-the-art of functional neuroanatomy of speech perception (Scott & Johnsrude, 2003; Hickok & Poeppel, 2000, 2004, 2007; Rauschecker & Scott, 2009). We briefly summarize the main consensus findings (Fig. 9.6A) that are based on functional neuroimaging (fMRI, positron emission tomography [PET], MEG/EEG) and lesion data.

Departing from the classical model in which both a posterior (Wernicke's) and an anterior (Broca's) area form the anatomic network, it is now argued that speech is processed in parallel in at least two streams, a ventral stream for speech-to-meaning mapping (a "what" stream), and a dorsal stream for speech-to-articulation mapping (a "how" stream). Both streams converge on prefrontal cortex, with a tendency for the ventral pathway to contact ventral prefrontal cortex (BA 44/45, also referred to as Broca's area), and the dorsal pathway to contact dorsal premotor regions (Hickok & Poeppel, 2007; Rauschecker & Scott, 2009). The dual path network operates both in a feedforward (bottom-up) and feedback (top-down) manner—highlighting, in turn, the need for algorithmic theories that have appropriate primitives to permit such bidirectional processing in real time. An additional feature of the inclusion in current models of both ventral (temporal–frontal) and dorsal (temporal–parietal–frontal) streams has been a renewed appreciation for the subtlety

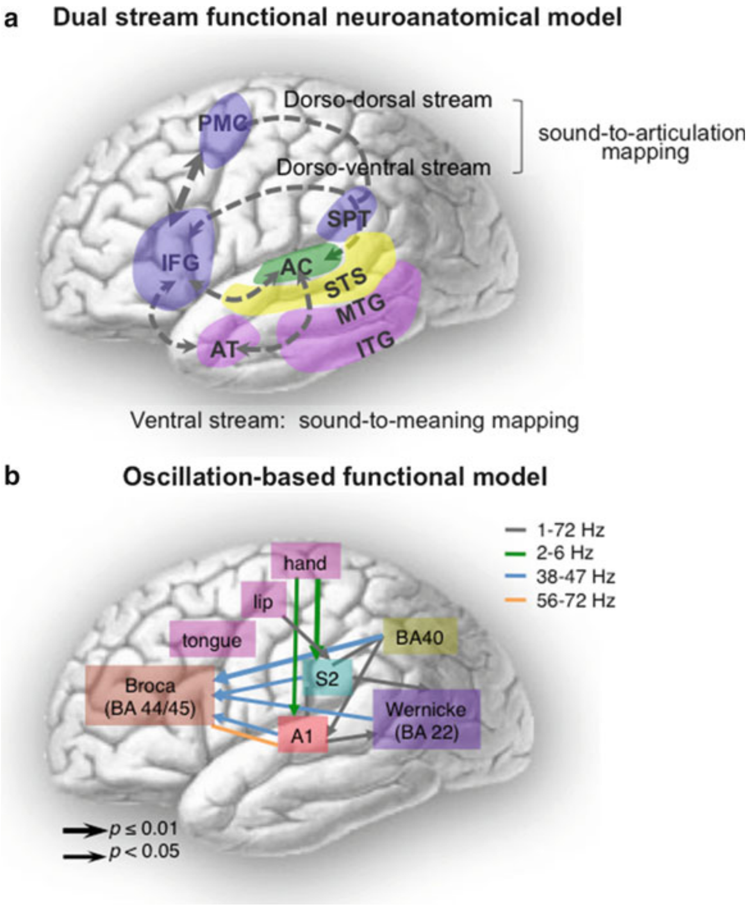


Fig. 9.6 Two functional neuroanatomical models of speech perception. (**a**) Model based on neuropsychology and functional neuroimaging data (PET and fMRI; after Hickok and Poeppel, 2007). (**b**) Model based on the propagation of resting oscillatory asymmetry during an audiovisual linguistic stimulation (a spoken movie). Modified from Giraud and Poeppel, 2012

of hemispheric specialization (Ueno et al., 2011). In particular, dorsal pathway structures (see Fig. 9.6A) appear much more strongly (left) lateralized, whereas the areas comprising the ventral processing stream(s), at least early on (e.g., superior temporal gyrus [STG], STS, medial temporal gyrus [MTG]), reveal robust bilateral contributions, whether assessed by hemodynamic or electrophysiological techniques. There is certainly no one-size-fits-all answer to hemispheric specialization for speech and language processing.

Historically, neuropsychological deficit-lesion research has been the main source of data regarding such anatomic models (Bates et al., 2003). In the context of dual stream proposals, the dorsal structures play a more central role in mediating output related computations. Because output tasks (e.g., word repetition) are the most frequently used instruments in clinical work to assess poststroke performance, there is thus a natural tendency to overemphasize the degree of left hemisphere dominance for speech and language. While output operations are apparently strongly lateralized to the dominant left hemisphere, the operations underlying comprehension are much more bilateral (Giraud et al., 2004). Various aspects of comprehension, including the recognition of voice, of prosody, and of components of lexical semantics have been strongly implicated as right-hemisphere functions. In sum, statements about speech and language lateralization must be taken with caution, requiring reference to the specific subroutines under consideration. For a related electrophysiological perspective on language comprehension, see the "PARLO" model (Federmeier, 2007), in which top-down predictive processing and production are argued to be left lateralized and more bottom-up processes right lateralized.

The functional anatomy corresponds to stages of perceptual analysis that are required for recognition: analysis of the acoustic signal; transformation to a phonetic or phonological code in order to link to stored linguistic information; contact with the stored representations, e.g., words; contact with the conceptual information linked to lexical entries; and in addition, depending on the tasks, retrieval of the articulatory code underlying spoken output; and combination of items to yield phrases, that is, compositional operations.

In human auditory cortex, the acoustic analysis of speech is initiated bilaterally in Heschl's gyrus. Although there are presumably qualitative differences in the type of processing that is carried out on each side (as outlined previously), metabolic and hemodynamic responses reveal no compelling asymmetries in the acoustic processing of speech sounds at the level visible to these techniques. A new meta-analysis on sublexical speech perception confirms that bilateral regions are fully involved in initial analyses, with subsequent mapping to phonology more left lateralized (Turkeltaub & Coslett, 2010). Depending on the task, phonological processing involves regions that are either anteroventral or posterodorsal to Heschl's gyrus along the superior temporal gyrus (BA22; Davis et al., 2005; Davis & Johnsrude, 2007). Passive listening and intelligibility tasks tend to involve anteroventral regions where there might be relatively stable phonological representations, possibly organized in a topographic manner (e.g., syllable or vowel maps; Obleser et al., 2006; Chang et al., 2010). Activation may extend to more anterior and ventral regions of the left temporal lobe. Which subroutines are executed in the more anterior ventral territories is a subject of intense current investigation, and proposals range from the anterior temporal lobe (ATL) mediating conceptual storage (Patterson et al., 2007) to linguistic combinatorics (Brennan et al., 2010). Phonetic-to-lexical mapping typically activates the STS and the MTG. The posterior third of middle temporal gyrus appears to play a key intermediate role in both recognizing and activating words in their formal, linguistic guise (STS to MTG mapping), as has been reviewed in Hickok and Poeppel (2007) and Lau et al. (2008). Further, speech production tasks implicate MTG in lexical representations before articulation (Indefrey & Levelt, 2004). Finally, there are reasons to believe that the meaning of words is activated, preactivated, or selected in MTG (for recent review, see Lau et al., 2008)

The extent to which STS and MTG activation is bilateral in the context of processing word form and word meaning is unresolved. A growing body of data suggests that here, too, the bilateral contribution has been underestimated. For example, in an fMRI study, blood oxygenation level–dependent (BOLD) responses to vocoded speech before and after subjects had learned to understand its linguistic content were recorded and clearly bilateral activation of the MTG (BA21) was observed. Giraud et al. (2004) concluded that it was essentially the early, phonological, steps of analysis that were more lateralized, but not the semantic analysis.

In contrast to identification or "what"-type tasks mediated by ventral stream temporal lobe regions, the dorsal stream structures of auditory cortex (as well as parietal and frontal lobes) play a more critical role in sensorimotor aspects of speech processing. However, there are conflicting hypotheses about the dorsal stream's contributions, ranging from (1) processing spectral changes over time ("how" pathway) to (2) extracting relevant sound features and matching them with stored templates of motor responses ("do" pathway) to (3) transforming auditory representations of speech into motor programs for speech gestures. The data motivating the differing research questions derive mostly from imaging studies and neuropsychological patient data. Electrophysiological experiments have, to date, contributed less to the discussions of concurrent processing streams and the differential role of dorsal structures. Two brain regions have lately received special attention, Spt (Sylvian parietotemporal) and intraparietal sulcus (IPS). Imaging experiments in which subjects are required to generate overt or covert articulated outputs typically activate regions that are posterior to Heschl's gyrus, the posterior planum temporale, and the supramarginal gyrus located just above Heschl's gyrus in the parietal operculum (Kell et al., 2010). The two latter regions merge in area Spt, which is argued to carry out the sensorimotor transformations underlying speech and other vocal tract activities (Hickok & Poeppel, 2007). A different line of research, explicitly testing feedforward and feedback auditory processing in audiovocal integration in musicians, has implicated the IPS (and its connectivity to STS) in the computations linking perception and production (Zarate & Zatorre, 2008; Zarate et al., 2010). This aspect of dorsal pathway function is briefly revisited in Section 9.4.3.

9.4.2 Broadening the Empirical Scope: an Oscillation-Based Functional Model

The chapter has emphasized a neurophysiological perspective, and especially the potential role of neuronal oscillations as "administrative mechanisms" to parse and decode speech signals. Does such a focus converge with the functional anatomic models mentioned above? Recent experimental research has begun addressing this issue directly and developed a functional anatomic model solely derived from recordings of neuronal oscillations. Based on analyses of the sources of oscillatory activity, that is, brain regions showing asymmetric theta/gamma activity at rest and under linguistic stimulation, Morillon et al. (2010) propose a new functional model of speech and language processing (Fig. 9.6B) that links elegantly to the textbook anatomy (illustrated in Fig. 9.6A). This model is grounded in a "core network" showing left oscillatory dominance at rest (no linguistic stimulation, no task), encompassing auditory, somatosensory, and motor cortices, and BA40 in inferior parietal cortex. The strongest asymmetries are observed in motor cortex and in BA40, which hence presumably play an important causal role in left hemispheric dominance during language processing. Critically, the proposed core network does not include Wernicke's (BA22) and Broca's (BA44/45) areas, despite the fact that both are classically related to speech and language processing. Interestingly, whereas these areas show no sign of asymmetry at rest, they "inherit" left dominant oscillatory activity during linguistic processing from the putative core regions. The model argues that posterior superior temporal cortex (Wernicke's area) inherits its profile from auditory and somatosensory cortices, while Broca's area inherits its profile from all posterior regions including auditory, somatosensory, Wernicke, and BA40. This model specifies that posterior regions share their oscillatory activity over the whole range of frequencies examined (1–72 Hz), while Broca's area inherits only the gamma range of the posterior oscillatory activity. This might reflect that oscillatory activity in Broca's area does not exclusively pertain to language. Finally, an important feature of the model is the influence of the motor lip and hand areas on auditory cortex oscillatory activity on the delta/theta scale, which underlines the importance of syllable and co-speech gesture production rates, on the receptive auditory sampling, and its asymmetric implementation. This model is compatible with a hardwired alignment of speech perception and production capacities at a syllable but not at a phonemic scale, suggesting that sensory/motor alignment at the phonemic scale is presumably acquired. Using an approach entirely driven by oscillations, this model is largely consistent with the previous one, but places a new emphasis on hardwired auditory–motor interactions, and on a determinant role of BA40 in language lateralization, which remains to be clarified.

9.4.3 The Role of the Auditory Cortex in Speech Production

Arguing that auditory cortex lies at the basis of speech perception is hardly surprising or insightful, yet it is worth remembering that the literature on speech recognition has been most deeply influenced by the *motor theory of speech perception* (Liberman et al., 1967; Corballis, 2009). This theory holds that listeners recover the intended articulatory gestures of the speaker, that is, properties of a motoric representation, and a substantial literature argues that motor cortical areas show activation in the relevant situations (Wilson et al., 2004; Pulvermueller et al., 2006). A different perspective can be characterized as the *sensory theory of speech production*. The idea is developed in some detail in Hickok et al. (2011). In the latter view, somatosensory (vocal tract configuration) and auditory (spectrotemporal) goals lie at the basis of speech production, from which claim it follows that *auditory cortex is centrally involved in production as well as perception*. In contrast, the causal role of motor cortical structures for perception is thereby challenged. Both models depicted in Figure 9.6 underscore that brain systems for speech perception and speech production are intimately linked at both functional and anatomical levels.

What is the hypothesized role of auditory cortical regions in production tasks? Both imaging (fMRI) and electrophysiological (MEG) studies suggest that speech decoding structures in human auditory cortex are preactivated during speech planning (e.g. Kell et al., 2010; Tian & Poeppel, 2010), presumably through input from premotor cortex as well as parietal areas. These areas provide feedback to the motor system for the control of speech production. The discussion surrounding the contribution of sensory areas such as auditory cortex to production is largely embedded in the framework of internal forward models. Such models have been elaborated in detail by Guenther and colleagues for production (Guenther, 2006; Guenther et al., 2006) and receive support in electrophysiological studies demonstrating the predictive aspect of production via efference copies (Eliades & Wang, 2008; Tian & Poeppel, 2010), and align well with large-scale psycholinguistic models of perception and production (Hickok et al., 2011).

9.4.4 A Predictive (Bayesian) View on Speech Processing

When processing continuous speech, as outlined in Section 9.3, the brain needs to simultaneously carry out acoustic and linguistic operations: at every instant there is both acoustic input to be processed and meaning to be calculated from the preceding input. Discretization using phases during which cortical neurons are either highly or weakly receptive to input is one computational principle that could ensure constant alternation between sampling the input and matching this input onto higher-level, more abstract representations. The Bayesian perspective on this issue assumes that the brain decodes sounds by constantly generating inferences about what is and will be said, on the basis of the quickest and crudest neural representation it can make with an acoustic input (Poeppel et al., 2008). Discretization at multiple timescales and Bayesian speech decoding principles are gathered in the conceptual model proposed in Figure 9.7 (adapted from Poeppel et al., 2008). In this model, neural representations of speech sounds are activated via both (1) a bottom-up process and (2) a higher-order prior based on previous input, knowledge of language, etc. These assumptions may correspond to coarse "preactivation" of representations, which subsequently accelerate the match between representation and input. Such priors can theoretically be formed at every representational level, acoustic, phonological, lexical, etc. Figure 9.7 illustrates, in three horizontal levels, the mapping from an acoustic input on the left to an output lexical item (or string of words) on the right. The boxes at the bottom exemplify putative types of analyses that are required for successful recognition. Something like these proposed analyses must be correct on logical grounds—and this chapter argues that the multitime-resolution analysis plays one helpful role in the overall process. The three boxes in the middle level make reference to which cortical areas are implicated for some of the operations.

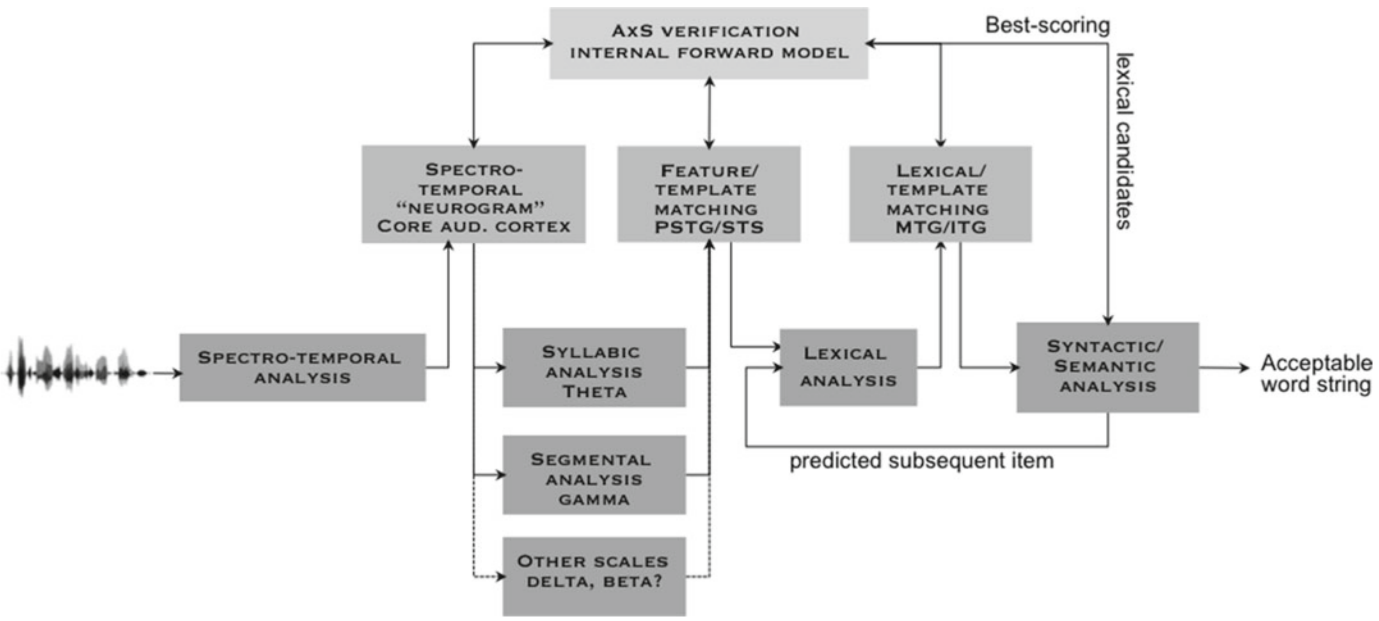


Fig. 9.7 Block-diagram of hypothesized operations taking place during speech perception within an analysis-by-synthesis framework (after Poeppel et al., 2008). The lower boxes represent different levels of sound to word mapping. The top box corresponds to a hypothetical level where an internal forward model in formed. The intermediate boxes show possible computations resulting from the interaction with bottom and top levels operations. The internal forward model is updated periodically with each new neuro-sample that is available (possibly every 30 ms and 200 ms). A detailed spectrotemporal analysis of the acoustic signal is available in bilateral primary auditory areas. Segmental (gamma) and syllabic-size (theta) samples are then formed close to auditory cortex, in the superior temporal gyrus (STG), and in the superior temporal sulcus (STS), respectively. The mapping between featural information and lexical entries occurs in the STS, and lexical processes in the medial temporal gyrus (MTG). Further compositional semantic and syntactic steps are carried out in prefrontal cortex. Top-down forward model signals reach many of the aforementioned steps/regions

Note that in this visualization, it is ventral stream areas that are principally implicated. The box on top identifies two of the putative types of “heuristics” or algorithms that are under consideration: the internal forward models mentioned previously, and analysis-by-synthesis (cf. Poeppel et al., 2008), an algorithm for perception suggested in the 1950s that takes small bits of input and generates, sequentially, the hypothesized output compatible with an input string, iteratively yielding better matches. On both of these concepts of processing, much of perception is actually achieved by a form of internal prediction and/or production, yet these models are rather different from motor theories. A proposal in very similar spirit to the one exemplified in Figure 9.7 is the “reverse hierarchy theory,” a conceptualization developed to meet certain challenges in visual object recognition (Hochstein & Ahissar, 2002) and recently extended to speech processing (Nahum et al., 2008).

In their experiment, which effectively illustrates the tension between bottom-up and top-down components of speech decoding, Giraud et al. (2004) contrasted functional brain images in which identical vocoded stimuli could be either understood or not depending on previous experience. Before exposure to the corresponding natural speech stimuli, participants perceived vocoded speech as noise, whereas after exposure they perceived it as speech and could reconstruct the meaning from degraded sounds. At a behavioral level this exemplifies that perceiving linguistic content in speech is not merely the result of acoustic processing. At the functional neuroimaging level, very little neural activation corresponds to speech comprehension per se; the essential part of the process corresponds to auditory search, which reflects iterative matching between hypothesis and incoming input.

It is difficult to characterize the neurophysiological processes underlying top-down control on speech processing using the auditory modality alone, precisely because top-down and bottom-up influences concurrently operate on the same neuronal target. Van Wassenhove et al. (2005) designed a study using natural audiovisual speech where it is the visual modality that primes the auditory modality (see van Wassenhove and Schroeder, Chapter 11). Because when we speak the onset of visual movements leads the auditory onset by about 150 ms, the brain can infer/predict auditory input from visual movements. Using this ecological audiovisual setting, it is possible to record with EEG or MEG in humans both the response to the visual input (the predictor) and the impact of visual prediction on auditory response to speech. Van Wassenhove et al. (2005) showed that the early auditory response is accelerated by visual input, with the degree of temporal facilitation related to how informative the facial configuration was. For example, seeing a speaker with the mouth in a bilabial configuration (i.e., poised to say /ba/ or /pa/ or /ma/) leads to up to 25 ms of facilitation because such a small set of auditory targets is possible. Using an identical setting, that is, videos of a speaker pronouncing syllables, Arnal et al. (2009, 2011) have refined this approach, showing that facilitation also involves a reduction in the amplitude of the response. Critically, both latency and amplitude reductions are proportional to the informational value contained in the visual input. Syllables starting with a bilabial consonant, for example, /pa/ /ma/, are more informative, hence more predictive, than when the consonant is formed at the back, for example, /ga/, /ka/. This shows that predictions made by the brain on the basis of rather crude sensory information strongly influence speech processing. Bayesian models of cortical responses stipulate that at each level of the hierarchy the neural response that is propagated forward reflects the difference between a prediction and the actual input (Friston, 2010). If correctly predicted, a stimulus therefore gives rise to a smaller cortical response than if unexpected. This phenomenon could be accounted for by the size to neuronal population that responds to a stimulus. When a speech stimulus is not predicted, the brain could respond with a large response reflecting the involvement of a broader neuronal population. This neural strategy, although ensuring that the brain does not miss a stimulus, is both cognitively costly and imprecise. As soon as a stimulus is either recognized or correctly anticipated the size of the recruited neuronal population drops, reflecting a more precise, focal activation in auditory cortex. Other accounts have recently been advanced for such phenomena (Wacongne et al., 2011).

9.5 Summary

This chapter engages, at the outset, some potential terminological confusion. "Speech perception" is many things to many people, and the failure to distinguish carefully between terms that have overlapping, obtuse, or no definitions has led to some unfortunate misunderstandings in the literature. Section 9.2 summarizes some of the salient properties of the speech signal that lie at the basis of what human auditory cortex must process. Particular emphasis is placed on some temporal attributes, including the low modulation frequencies in speech that play a special role for intelligibility. The section covers the sensitivity to frequency and the sensitivity to time of cortical neurons. In addition, the high degree of tuning to spectrotemporal modulation is discussed. In Section 9.3, the chapter turns to the processing of speech as a continuous signal. One of the central challenges is here called the discretization problem: how does auditory cortex create chunks of the appropriate temporal granularity for further computation? The solution that is pursued in this chapter builds on the concept of neuronal oscillations. In particular, oscillations in multiple frequencies (theta, gamma) are argued to provide the right mechanisms to align with the speech signal and sample the speech signal at different rates. Multitime-resolution processing and the asymmetric sampling in time (AST) hypothesis are summarized. Section 9.4 outlines large-scale models. First, the consensus functional anatomic models are discussed. The dual stream model is highlighted, and new neuro-oscillatory data are reviewed that extend and strengthen such a multiple pathway approach to speech perception. Further, it is highlighted that auditory cortex plays a critical role in speech production, reversing the standard roles in the literature that emphasize the role of motor cortex and speech perception. In terms of functional analysis, the notion of an internal forward model is presented, building on the observation that much of perceptual analysis has a strong predictive component. Finally, audiovisual speech experiments are shown to test some of the predictions of these models.

A comprehensive and explanatory neurocognitive model of speech perception remains an ambitious goal. It is worth remembering that speech perception is a task that is executed with automaticity and great ease by even early learners, but that is handled surprisingly poorly by even the most sophisticated automatic devices. The brain appears to solve this very challenging problem by breaking it down into parts: it is broken down in space, by implementing the functional anatomy as multiple concurrent streams, and it is broken down in time, by implementing multitime-resolution mechanisms that analyze information on multiple scales concurrently. Like all models, surely the ones presented here are dramatically underspecified and will turn out to be naïve. That being said, one hopes that they are wrong in an interesting way, leading to new research questions and incremental progress on this foundational question about human perception.

Acknowledgments The preparation of this manuscript was supported by CNRS to A. L. G. and NIH 2R01DC05660 to D. P.

References

Abeles, M. (1982). Role of the cortical neuron: Integrator or coincidence detector? *Israel Journal of Medical Sciences*, 18(1), 83–92.

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences of the USA*, 98(23), 13367–13372.

Allen, J. B. (2005). Articulation and intelligibility. *Synthesis Lectures on Speech and Audio Processing*, 1(1), 1–124.

Arnal, L. H., Morillon, B., Kell, C. A., & Giraud, A. L. (2009). Dual neural routing of visual facilitation in speech processing. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 29(43), 13445–13453.

Arnal, L. H., Wyart, V., & Giraud, A.L. (2011). Transitions in neural oscillations reflect prediction errors generated in audiovisual speech, *Nature Neuroscience*, 16(6), 794-801.

Atencio, C. A., Sharpee, T. O., & Schreiner, C. E. (2009). Hierarchical computation in the canonical auditory cortical circuit. *Proceedings of the National Academy of Sciences of the USA*, 106(51), 21894–21899.

Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., & Dronkers, N. F. (2003). Voxel-based lesion-symptom mapping. *Nature Neuroscience*, 6(5), 448–450.

Bendor, D., & Wang, X. (2006). Cortical representations of pitch in monkeys and humans. *Current Opinion in Neurobiology*, 16(4), 391–399.

Bendor, D., & Wang, X. (2007) Differential neural coding of acoustic flutter within primate auditory cortex. *Nature Neuroscience*, 10(6), 763–771.

Binzegger, T., Douglas, R. J., & Martin, K. A. (2007). Stereotypical bouton clustering of individual neurons in cat primary visual cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 27(45), 12242–12254.

Blumstein, S. E., Myers, E. B., & Rissman, J. (2005). The perception of voice onset time: An fmri investigation of phonetic category structure. *Journal of Cognitive Neuroscience*, 17(9), 1353–1366.

Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience*, 8(3), 389–395.

Borgers, C. & Kopell, N. J. (2008). Gamma oscillations and stimulus selection. *Neural Computations*, 20(2), 383–414.

Borgers, C., Epstein, S., & Kopell, N. J. (2005). Background gamma rhythmicity and attention in cortical local circuits: A computational study. *Proceedings of the National Academy of Sciences of the USA*, 102(19), 7002–7007.

Brennan, J., Nir, Y., Hasson, U., Malach, R., Heeger, D. J., & Pylkkänen, L. (2010). Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain and Language*. doi:10.1016/j.bandl.2010.04.002

Britvina, T., & Eggermont, J. J. (2007). A Markov model for interspike interval distributions of auditory cortical neurons that do not show periodic firings. *Biological Cybernetics*, 96(2), 245–264.

Brugge, J. F., Nourski, K. V., Oya, H., Reale, R. A., Kawasaki, H., Steinschneider, M., & Howard, M. A., 3rd (2009). Coding of repetitive transients by auditory cortex on Heschl's gyrus. *Journal of Neurophysiology*, 102(4), 2358–2374.

Canolty, R. T., & Knight, R. T. (2010). The functional role of cross-frequency coupling. *Trends in Cognitive Sciences*, 14(11), 506–515.

Chait, M., Poeppel, D., & Simon, J. Z. (2006). Neural response correlates of detection of monaurally and binaurally created pitches in humans. *Cerebral Cortex*,, 16(6), 835–848.

Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., & Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience*, 13(11), 1428–1432.

Cleary, M., & Pisoni, D. B. (2001). Speech perception and spoken word recognition: Research and theory. In B. Goldstein (Ed.), *Handbook of perception* (pp. 499–534). Cambridge, MA: Blackwell.

Corballis, M. C. (2009). The evolution of language. *Annals of the New York Academy of Sciences*, 1156, 19–43.

da Costa, N. M., & Martin, K. A. C. (2010). Whose cortical column would that be? *Frontiers in Neuroanatomy*. doi: 10.3389/fnana.2010.00016.

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research*, 229(1–2), 132–147.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal Experimental Psychology General*, 134(2), 222–241.

Ding, N., & Simon, J. Z. (2009). Neural representations of complex temporal modulations in the human auditory cortex. *Journal of Neurophysiology*, 102(5), 2731–2743. Eger, E., Michel, V., Thirion, B., Amadon, A., Dehaene, S., & Kleinschmidt, A. (2009). Deciphering cortical number coding from human brain activity patterns. *Current Biology*, 19(19), 1608–1615.

Elhilali, M., J. B. Fritz, Klein, D. J., Simon, J. Z., & Shamma, S. A. (2004). Dynamics of precise spike timing in primary auditory cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 24(5), 1159–1172.

Eliades, S. J., & Wang, X. (2008). Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature*, 453(7198), 1102–1106.

Elliott, T. M., & Theunissen, F. E. (2009). The modulation transfer function for speech intelligibility. *PLoS Computational Biology*, 5(3), e1000302. doi:10.1371/journal.pcbi.1000302

Faulkner, A., Rosen, S., & Smith, C. (2000). Effects of the salience of pitch and periodicity information on the intelligibility of four-channel vocoded speech: Implications for cochlear implants. *The Journal of the Acoustical Society of America*, 108(4), 1877–1887.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, 44(4), 491–505.

Formisano, E., Kim, D. S., Di Salle, F., van de Moortele, P. F., Ugurbil, K., & Goebel, R. (2003). Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron*, 40(4), 859–869.

Formisano, E., De Martino, F., Bonte, M., & Goebel, R. (2008). "Who" is saying "what"? Brain-based decoding of human voice and speech. *Science*, 322(5903), 970–973.

Friederici, A. D., Kotz, S. A., Scott, S. K., & Obleser, J. (2010). Disentangling syntax and intelligibility in auditory language comprehension. *Human Brain Mapping*, 31(3), 448–457.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138.

Gaese, B. H., & Ostwald, J. (1995). Temporal coding of amplitude and frequency modulation in the rat auditory cortex. *The European Journal of Neuroscience*, 7(3), 438–450.

Gaskell, M. G., & Marslen-Wilson, W. D. (2002). Representation and competition in the perception of spoken words. *Cognitive Psychology*, 45(2), 220–266.

Ghitza, O. (2011). Linking speech perception and neurophysiology: Speech decoding guided by cascaded oscillations locked to the input rhythm. *Frontiers in Psychology*, 2:130.

Ghitza, O., & Greenberg, S. (2009). On the possible role of brain rhythms in speech perception: Intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica*, 66(1–2), 113–126.

Giraud, A.L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, in press.

Giraud, A. L., & Price, C. J. (2001). The constraints functional neuroimaging places on classical models of auditory word processing. *Journal of Cognitive Neuroscience*, 13(6), 754–765.

Giraud, A. Lorenzi, C., Ashburner, J., Wable, J., Johnsrude, I., Frackowiak, R., & Kleinschmidt, A. (2000). Representation of the temporal envelope of sounds in the human brain. *Journal of Neurophysiology*, 84(3), 1588–1598.

Giraud, A. L., Kell, C., Thierfelder, C., Sterzer, P., Russ, M. O., Preibisch, C., & Kleinschmidt, A. (2004). Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. *Cerebral Cortex*, 14(3), 247–255.

Giraud, A. L., Kleinschmidt, A., Poeppel, D., Lund, T. E., Frackowiak, R. S., & Laufs, H. (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron*, 56(6), 1127–1134.

Greenberg, S., & Ainsworth, W. A. (2006). *Listening to speech: An auditory perspective*. Mahwah, NJ: Lawrence Erlbaum.

Greenberg, S. & Arai, T. (2001). The relation between speech intelligibility and the complex modulation spectrum. *Proceedings of the 7th Eurospeech Conference on Speech Communication and Technology (Eurospeech-2001)*, 473-476.

Greenberg, S., & Kingsbury, B. E. D. (1997). The modulation spectrogram: In pursuit of an invariant representation of speech. *Proceedings of the 1997 IEEE International Conference on Acoustics*, *Speech*, *and Signal Processing (ICASSP '97)-Volume 3*

Griffiths, T. D., Kumar, S., Sedley, W., Nourski, K. V., Kawasaki, H., Oya, H., et al. (2010) Direct recordings of pitch responses from human auditory cortex. *Current Biology*, 20(12), 1128–1132.

Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders*, 39(5), 350–365.

Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, 96(3), 280–301.

Harnad, S. R. (1987). *Categorical perception: The groundwork of cognition*. Cambridge, UK: Cambridge University Press.

Hawkins, S. (1999). Reevaluating assumptions about speech perception: Interactive and integrative theories. In J. M. Pickett (Ed.), *The acoustics of speech communication* (pp. 232–288). Boston: Allyn and Bacon.

Heil, P. (1997a). Auditory cortical onset responses revisited. I. First-spike timing. *Journal of Neurophysiology*, 77(5), 2616–2641.

Heil, P. (1997b). Auditory cortical onset responses revisited. II. Response strength. *Journal of Neurophysiology*, 77(5), 2642–2660.

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, 4(4), 131–138.

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, 92(1–2), 67–99.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393–402.

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron*, 69(3), 407–422. Hochstein, S., &

Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, 36(5), 791–804.

Holcombe, A. O. (2009). Seeing slow and seeing fast: Two limits on perception. *Trends in Cognitive Sciences*, 13(5), 216–221.

Hromádka, T., & Zador, A. M. (2009). Representations in auditory cortex. *Current Opinion in Neurobiology*, 19(4), 430–433.

Hutsler, J., & Galuske, R. A. (2003). Hemispheric asymmetries in cerebral cortical networks. *Trends in Neurosciences*, 26(8), 429–435.

Indefrey, P., & Levelt, W. J. (2004) The spatial and temporal signatures of word production components. *Cognition*, 92(1–2), 101–144.

Jamison, H. L., Watkins, K. E., Bishop, D. V., & Matthews, P. M. (2006). Hemispheric specialization for processing auditory nonspeech stimuli. *Cerebral Cortex*,, 16(9), 1266–1275.

Joris, P. X., Schreiner, C. E., & Rees, A. (2004). Neural processing of amplitude-modulated sounds. *Physiological Reviews*, 84(2), 541–577.

Kanedera, N., Arai, T., Hermansky, H., & Pavel, M. (1999). On the relative importance of various components of the modulation spectrum for automatic speech recognition. *Speech Communication*, 28(1), 43–55.

Kayser, C., Logothetis, N. K., & Panzeri, S. (2010). Millisecond encoding precision of auditory cortex neurons. *Proceedings of the National Academy of Sciences of the USA*, 107(39), 16976–16981.

Kell, C. A., Morillon, B., Kouneiher, F., & Giraud, A. L. (2010). Lateralization of speech production starts in sensory cortices—a possible sensory origin of cerebral left dominance for speech. *Cerebral Cortex*,. doi:10.1093/cercor/bhq167

Klatt, D. H. (1989). Review of selected models of speech perception. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 169–226). Cambridge, MA: MIT Press.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience*, 9(12), 920–933.

Laver, J. (1994). *Principles of phonetics. Cambridge textbooks in linguistics*. New York: Cambridge University Press.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74(6), 431–461.

Loebach, J. L., & Wickesberg, R. E. (2008). The psychoacoustics of noise vocoded speech: A physiological means to a perceptual end. *Hearing Research*, 241(1–2), 87–96.

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, 54(6), 1001–1010.

Luo, H., Wang, Y., Poeppel, D., & Simon, J. Z. (2006). Concurrent encoding of frequency and amplitude modulation in human auditory cortex: MEG evidence. *Journal of Neurophysiology*, 96(5), 2712–2723.

Luo, H., Boemio, A., Gordon, M., & Poeppel, D. (2007a). The perception of FM sweeps by Chinese and English listeners. *Hearing Research*, 224(1–2), 75–83.

Luo, H., Wang, Y., Poeppel, D., & Simon, J. Z. (2007b). Concurrent encoding of frequency and amplitude modulation in human auditory cortex: Encoding transition. *Journal of Neurophysiology*, 98(6), 3473–3485.

Mantini, D., Perrucci, M. G., Del Gratta, C., Romani, G. L., &, Corbetta, M. (2007). Electrophysiological signatures of resting state networks in the human brain. *Proceedings of the National Academy of Sciences of the USA*, 104(32), 13170–13175.

McClelland, J. L., & Elman, J. L. (1986). The trace model of speech perception. *Cognitive Psychology*, 18(1), 1–86.

Middlebrooks, J. C. (2008). Auditory cortex phase locking to amplitude-modulated cochlear implant pulse trains. *Journal of Neurophysiology*, 100(1), 76–91.

Miller, G. A. (1951). *Language and communication*. New York: McGraw-Hill.

Monahan, P. J., & Idsardi, W. J. (2010). Auditory sensitivity to formant ratios: Toward an account of vowel normalization. *Language and Cognitive Processes*, 25(6), 808–839.

Morillon, B., Lehongre, K., Frackowiak, R. S., Ducorps, A., Kleinschmidt, A., Poeppel, D., & Giraud, A. L. (2010). Neurophysiological origin of human brain asymmetry for speech and language. *Proceedings of the National Academy of Sciences of the USA*, 107(43), 18688–18693.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385(6615), 432–434.

Nahum, M., Nelken, I., & Ahissar, M. (2008). Low-Level information and high-level perception: The case of speech in noise. *PLoS Biology*, 6(5), e126.

Nelken, I., Bizley, J. K., Nodal, F. R., Ahmed, B., King, A. J., & Schnupp, J. W. (2008). Responses of auditory cortex to complex stimuli: Functional organization revealed using intrinsic optical signals. *Journal of Neurophysiology*, 99(4), 1928–1941.

Obleser, J., Boecker, H., Drzezga, A., Haslinger, B., Hennenlotter, A., Roettinger, M., et al. (2006). Vowel sound extraction in anterior superior temporal cortex. *Human Brain Mapping*, 27(7), 562–571.

Obleser, J., Eisner, F., & Kotz, S. A. (2008). Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 28(32), 8116–8123.

Overath, T., Kumar, S., von Kriegstein, K., & Griffiths, T. D. (2008). Encoding of spectral correlation over time in auditory cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 28(49), 13268–13273.

Pardo, J. S., & Remez, R. E. (2006). The perception of speech. In M. Traxler & M. A. Gernsbacher (Eds.), *The handbook of psycholinguistics*, 2nd ed. (pp. 201–248). New York: Academic Press.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, 8(12), 976–987.

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, 36(4), 767–776.

Petkov, C. I., Kayser, C., Augath, M., & Logothetis, N. K. (2006). Functional imaging reveals numerous fields in the monkey auditory cortex. *PLoS Biology*, 4(7), e215.

Phillips, D. P., Hall, S. E., & Boehnke, S. E. (2002). Central auditory onset responses, and temporal asymmetries in auditory perception. *Hearing Research*, 167(1–2), 192–205.

Pickett, J. M. (1999). *The acoustics of speech communication*. Boston: Allyn and Bacon.

Pienkowski, M., & Eggermont, J. J. (2010). Nonlinear cross-frequency interactions in primary auditory cortex spectrotemporal receptive fields: A Wiener-Volterra analysis. *Journal of Computational Neuroscience*, 28(2), 285–303.

Poeppel, D. (2001). Pure word deafness and the bilateral processing of the speech code. *Cognitive Science*, 25(5), 679–693.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as asymmetric sampling in time. *Speech Communication*, 41(1), 245–255.

Poeppel, D., & Monahan, P. J. (2008). Speech perception: Cognitive foundations and cortical implementation. *Current Directions in Psychological Science*, 17(2), 80.

Poeppel, D., Idsardi, W. J., & van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 363(1493), 1071–1086.

Pöppel, E. (1988). *Mindworks: Time and conscious experience*. Boston: Harcourt Brace Jovanovich.

Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences of the USA*, 103(20), 7865–7870.

Rabiner, L., & Juang, B. H. (1993). Fundamentals of speech recognition. Englewood Cliffs, NJ:Prentice-Hall

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6), 718–724.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science*, 212(4497), 947–949.

Roberts, B., Summers, R. J., & Bailey, P. J. (2011). The intelligibility of noise-vocoded speech: Spectral information available from across-channel comparison of amplitude envelopes. *Proc. R. Soc. B*, 278(1711), 1595–1600.

Rosen, S. (1992). Temporal information in speech: Acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London B:*, *Biological Sciences*, 336(1278), 367–373.

Saoud, H., Josse, G., Bertasi, E., Truy, E., Chait, M., & Giraud, A-L. (2012). Brain-speech alignment enhances auditory cortical responses and pseech perception. *The Journal of Neuroscience*, in press.

Saberi, K., & Perrott, D. R. (1999). Cognitive restoration of reversed speech. *Nature*, 398(6730), 760.

Schönwiesner, M., & Zatorre, R. J. (2009). Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fmri. *Proceedings of the National Academy of Sciences of the USA*, 106(34), 14611–14616.

Schroeder, C. E., & Lakatos, P. (2009a). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, 32(1), 9–18.

Schroeder, C. E., & Lakatos, P. (2009b). The gamma oscillation: Master or slave? *Brain Topography*, 22(1), 24–26.

Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*, 26(2), 100–107.

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain: A Journal of Neurology*, 123(Pt 12), 2400–2406.

Scott, S. K., Rosen, S., Lang, H., & Wise, R. J. (2006). Neural correlates of intelligibility in speech investigated with noise vocoded speech-a positron emission tomography study. *The Journal of the Acoustical Society of America*, 120(2), 1075–1083.

Shamir, M., Ghitza, O., Epstein, S., & Kopell, N. (2009). Representation of time-varying stimuli by a network exhibiting oscillations on a faster time scale. *PLoS Computational Biology*, 5(5), e1000370.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, 270(5234), 303.

Sharma, A., & Dorman, M. F. (1999).Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *The Journal of the Acoustical Society of America*, 106, 1078–1083.

Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, 416(6876), 87–90.

Souza, P., & Rosen S. (2009). Effects of envelope bandwidth on the intelligibility of sine- and noise-vocoded speech. *The Journal of the Acoustical Society of America*, 126(2), 792–805.

Steeneken, H. J. M., & Houtgast, T. (1980). A physical method for measuring speech-transmission quality. *The Journal of the Acoustical Society of America*, 67(1), 318–326.

Stevens, K. N. (1998). *Acoustic phonetics*. Cambridge, MA: MIT Press.

Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *The Journal of the Acoustical Society of America*, 111(4), 1872–1891.

Telkemeyer, S., Rossi, S., Koch, S. P., Nierhaus, T., Steinbrink, J., Poeppel, D., & Wartenburger, I. (2009). Sensitivity of newborn auditory cortex to the temporal structure of sounds. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 29(47), 14726–14733.

Tian, X., & Poeppel, D. (2010). Mental imagery of speech and movement implicates the dynamics of internal forward models. *Frontiers in Psychology*. doi: 10.3389/fpsyg.2010.00166

Tiesinga, P., & Sejnowski, T. J. (2009). Cortical enlightenment: Are attentional gamma oscillations driven by ING or PING? *Neuron*, 63(6), 727–732.

Turkeltaub, P. E., & Coslett, H. B. (2010). Localization of sublexical speech perception components. *Brain and Language*, 114(1), 1–15.

Ueno, T., Saito, S., Rogers, T. T., & Lambon-Ralph, M. A. (2011) Lichtheim 2: synthesizing aphasia and the neural basis of language in a neurocomputational model of the dual dorsal-ventral language pathways. *Neuron* 72: 385–96.

Van Rullen, R., & Koch, C. (2003). Is perception discrete or continuous? *Trends in Cognitive Sciences*, 7(5), 207–213.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the USA*, 102(4), 1181–1186.

von Kriegstein, K., Smith, D. R., Patterson, R. D., Kiebel, S. J., & Griffiths, T. D. (2010) How the human brain recognizes speech in the context of changing speakers. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 30(2), 629–638.

Wacongne, C., Labyt, E., van Wassenhove, V., Bekinchtein, T., Naccache, L., & Dehaene, S. (2011). Proceedings of the National Academy of Sciences, 108: 20754–9.

Wang, X. (2007). Neural coding strategies in auditory cortex. *Hearing Research*, 229(1–2), 81–93.

Wang, X. J. (2010). Neurophysiological and computational principles of cortical rhythms in cognition. *Physiological Reviews*, 90(3), 1195–1268.

Warrier, C., Wong, P., Penhune, V., Zatorre, R., Parrish, T., Abrams, D., & Kraus, N. (2009). Relating structure to function: Heschl's gyrus and acoustic processing. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 29(1), 61–69.

Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, 7(7), 701–702.

Womelsdorf, T., Schoffelen, J. M., Oostenveld, R., Singer, W., Desimone, R., Engel, A. K., & Fries, P. (2007). Modulation of neuronal interactions through neuronal synchronization. *Science* 316(5831), 1609–1612.

Zaehle, T., Wüstenberg, T., Meyer, M., & Jäncke, L. (2004). Evidence for rapid auditory perception as the foundation of speech processing: A sparse temporal sampling fmri study. *The European Journal of Neuroscience*, 20(9), 2447–2456.

Zarate, J. M., & Zatorre, R. J. (2008). Experience-dependent neural substrates involved in vocal pitch regulation during singing. *NeuroImage*, 40(4), 1871–1887.

Zarate, J. M., Wood, S., & Zatorre, R. J. (2010). Neural networks involved in voluntary and involuntary vocal pitch regulation in experienced singers. *Neuropsychologia*, 48(2), 607–618.

Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: Moving beyond the dichotomies. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 363(1493), 1087–1104.

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences*, 6(1), 37–46.

Chapter 10
Cortical Processing of Music

Robert J. Zatorre and Jean Mary Zarate

10.1 Introduction

Here's a commonplace experience: you are walking in a shopping mall when you hear a tune being played in the background. It takes you a moment but then you realize that it is a song that you last heard 20 years ago, which has now been redone—perhaps unfortunately—as an advertising jingle. Although the aesthetic experience associated with this little vignette may not be high, the ease with which our nervous system can carry out this kind of analysis belies the complexity involved. Consider: the music you hear is embedded in a background of irrelevant noise, so you need first to strip it away; you recognize the pattern of sound as the tune you are familiar with, even though none of the actual elements reaching your ear are the same as what you had originally encoded—the tempo, musical key, and instrument timbres may all be different; if the song has lyrics you must also separate the tonal component from the speech component to process each of them; the experience may also lead to retrieval of memories associated with the song; you could also begin to sing along with it, which means you must convert the information contained in the sound waves you hear to a set of motor commands that will produce similar sound waves from your vocal musculature; finally the song may lead you to experience emotion, which could range from annoyance to pleasure. The mechanisms that allow this complex cognitive chain of events to occur are far from being fully understood. This chapter aims to give readers an overview of what is known about

R.J. Zatorre (✉)
Cognitive Neuroscience Unit, Montréal Neurological Institute, McGill University,
3801 reu University, Montréal, Québec H3A 2B4, Canada
e-mail: robert.zatorre@mcgill.ca

J.M. Zarate
Department of Psychology, New York University, 6 Washington Place,
New York, NY 10003, USA
e-mail: jean.m.zarate@nyu.edu

D. Poeppel et al. (eds.), *The Human Auditory Cortex*, Springer Handbook of Auditory Research 43, DOI 10.1007/978-1-4614-2314-0_10,

the role of auditory cortex in processing and production of musical sounds, and an indication of the many open questions that remain. Understanding the neural and cognitive mechanisms involved in tonal and musical processes will yield insights into fundamental aspects of neural organization and function that would otherwise be difficult to obtain.

10.2 Pitch and Rhythm versus Speech: Building Blocks of Musical Processing

It is safe to say that music as we know it could not exist if our nervous system were not capable of some fairly sophisticated capabilities to process pitch. Pitch processing is also central to speech, particularly in tone languages (which are spoken by a majority of the world's citizens), wherein pitch contours serve to distinguish one word from another (review: Zatorre & Gandour, 2007). Pitch contours are also pertinent in non-tone languages, where they play both a syntactic as well as a paralinguistic role. But there are several critical distinctions between the use of pitch in music and speech that bear special mention. The use of pitch in musical contexts almost always involves a precise interval relationship between pitches—that is, a certain frequency ratio is used to define a melody or a chord; this is not the case in applications of pitch in speech, where the contour alone is used. In other words, when using pitch in a linguistic context, it is primarily the trajectory (rising/falling) that is relevant, whereas in music both the contour and the specific pitch intervals are critical (Dowling & Harwood, 1986). One way to think of this is that whereas it is easy to spot somebody singing a song with one "off-key" note, there is no true equivalent of this phenomenon in pitch used in speech. We return to this precision of pitch organization later. Another aspect of pitch organization that appears to be unique to music is harmony. The simultaneous playing of multiple tones creates chords, which in turn have complex relationships to one another in musical systems that employ tonality. Whereas chords, and the harmonic relationships engendered by them, are common in many musical systems, there is again no equivalent in speech; you do not coordinate the pitch of your speech to have a specific relationship to that of another concurrent speech stream. Indeed, multiple talkers generally generate confusion if they speak simultaneously, whereas coordinated simultaneous sounds are commonplace in most music.

The other element that is critical in all musical systems is rhythm, which refers to aspects of the temporal organization of unfolding sounds. A detailed discussion of the complexities of rhythm is beyond the scope of this chapter, but suffice it to say that there is a commonality in how temporal organization applies to speech and to music, but also some important differences. Perhaps the most salient of these differences is that in music there is typically a hierarchy of temporal organization that results in a metrical structure. Regular metrical structure is a common feature of music from many cultures and may be thought of in terms of a hierarchical framework of stronger and weaker events (perceived beats) that is inferred from the sound pattern, and unfolds over equal units of time. When listening to music, most people (even those without musical training), are able to extract this periodic higher-order organization, which allows for temporal expectancies. This is the basis for most people's ability to tap to the beat of a melody (Large & Palmer, 2002). Note that there is no equivalent to this effect in speech; speech has its own local temporal organization, but the concept of tapping to the beat of, say, somebody reading a news report, is nonsensical. For a detailed and lucid discussion of the many interesting parallels and differences between music and speech, the reader is referred to the volume by Patel (2008).

10.3 Neural Substrates of Basic Aspects of Pitch

We refer to primary and higher-order auditory cortex throughout this chapter. Human primary auditory cortex (see Clarke and Morosan, Chapter 2) is found typically in the medial portion of the first anterior transverse temporal gyrus, also known as Heschl's gyrus (HG). Galaburda and Sanides (1980) found higher-order auditory areas in surrounding portions of planum polare, anteriorly situated to HG, and planum temporale located posteriorly to HG (Galaburda & Sanides, 1980; Rivier & Clarke, 1997). For an in-depth discussion of the distribution of these auditory cortical regions, the reader should consult Clarke and Morosan (Chapter 2).

To understand the mechanisms involved with pitch processing, consider how the percept of pitch emerges from the physics of vibrating objects (for a review of these phenomena, see McDermott & Oxenham, 2008). Most sounds that result in perception of pitch contain periodicity; that is, the vibrations involved repeat at regular intervals. In many cases, and particularly those most relevant to music, these vibrations are related to one another by simple integer ratios. For instance, if you pluck a string, it will vibrate in its entire length, giving rise to one frequency, called the fundamental; it will also vibrate in halves, thirds, and so forth, and each of these give rise to a separate vibration; these are known collectively as harmonics. To simplify somewhat, we may say that the dominant pitch percept usually corresponds to the fundamental frequency. However, the harmonics play an important role, both in the percept of pitch itself, as well as in the tonal quality, or timbre, of the sound. Importantly, the pitch one perceives typically corresponds to the fundamental frequency, even if this particular component of the sound is weak, masked by other sounds, or absent. This phenomenon, often referred to as the "missing fundamental," may be thought of as a mechanism by which the nervous system "fills in" for missing information (Fig. 10.1). Another way to describe this phenomenon is in the context of perceptual constancy, which refers to the fact that we tend to perceive events as being relatively invariant despite significant changes in the actual stimulus energy that reaches our senses.

This missing-fundamental phenomenon has been the source of much experimentation, and has also been used to probe the neural substrates of pitch processing. One study tested missing-fundamental perception in human patients who had undergone

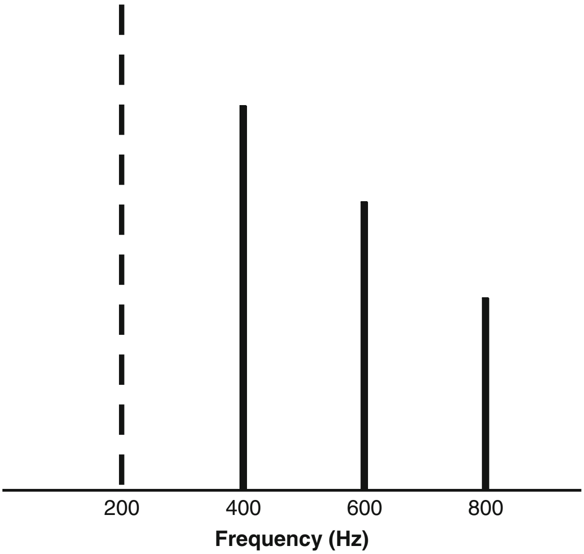


Fig. 10.1 An illustration of the "missing-fundamental" phenomenon. A presented set of harmonics at 400, 600, and 800 Hz creates the pitch percept of a fundamental frequency at 200 Hz, even if it is absent (represented by the dashed line). In a natural environment, those particular frequencies—all multiples of 200 Hz—would most likely have been produced by a sound source with a fundamental at that frequency

excision within specific subregions of auditory cortex (Zatorre, 1988), and found that the most severe deficit in perceiving the pitch occurred after lesions of the right lateral HG, and not with more anterior lesions sparing this area, nor with lesions on the left side. Importantly, such patients had no difficulty performing the task when the fundamental frequency was present, indicating the specificity of the effect. Thus, this study suggested specificity both at the level of lesion site as well as lesion side. The hemispheric difference suggested by this study was supported in a magnetoencephalography (MEG) experiment that showed that the dynamic pattern of brain activity recorded from right auditory cortex to missing-fundamental stimuli differed in persons who perceived the missing fundamental, whereas there was no difference in those who did not perceive it (Patel & Balaban, 2001); thus, it appears that right auditory cortex contains neural mechanisms allowing distinct missing fundamental tones to be differentiated from one another.

More recently, converging evidence for the importance of a region lateral to primary auditory cortex in processing missing fundamental pitch has been obtained from neurophysiological data in marmoset monkeys. Bendor and Wang showed that in a roughly homologous location to that identified in the human studies, there exist neurons that respond in an invariant manner to sounds that have very different harmonic structure, but that all share the same (absent) fundamental frequency (Bendor & Wang, 2005). These findings have led to the proposal that a pitch-sensitive region may exist, located lateral to primary auditory regions (Bendor & Wang, 2006), and that it may play a role in integrating information coming from multiple harmonics (presumably computed in earlier structures; cf. Hall and Garcia, Chapter 7 for further discussion).

Evidence for the concept of a pitch-sensitive area comes from a number of additional sources (Fig. 10.2). Lesion studies in humans not only point to the importance of a right lateral HG region for missing fundamental pitch, but also for another basic pitch-based skill: the ability to order two different pitches in terms of their direction (up or down). Patients with such lesions on the right side have a much larger threshold to perform this task than controls, or than those with equivalent left-auditory cortex lesions (Fig. 10.2A; Johnsrude et al., 2000). Importantly, the deficit lies not in the ability to perform the task as such, but rather in the size of the pitch interval required; that is, lesions of the right HG do not result in the abolition of pitch perception, but only in an inability to perceive fine pitch differences, suggesting that the right lateral HG region is important in terms of spectral resolution.

Turning to functional neuroimaging studies, we again find relatively consistent indications of heightened pitch sensitivity in a similar region. In several functional magnetic resonance imaging (fMRI) studies, lateral HG bilaterally was more active when a noise stimulus resulted in a perceived pitch (using a stimulus known as iterated ripple noise) than when the noise does not carry pitch information (Griffiths et al., 1998; Patterson et al., 2002). Similarly, Penagos et al. (2004) demonstrated that an area in a comparable location of lateral HG was sensitive to the perceived pitch salience of the stimulus. Although some fMRI studies have suggested that this pitch-sensitive region may be confined to only certain types of stimuli (Hall & Plack, 2009), other studies have found consistent evidence for recruitment of the same lateral HG region to several distinct pitch-producing stimuli (Puschmann et al., 2010), leading to the suggestion that it is the percept of pitch which is associated with activity in this region, and not any specific stimulus feature. In addition, Hyde et al. (2008) found that fMRI signal increased systematically in a region lateral and posterior to the right HG as the size of the pitch interval in a pattern increased (Fig. 10.2B). The asymmetry in this study was relative and not absolute, in that a region in the left lateral HG did respond, but only when the size of the pitch change was sufficiently large; this finding is in accord with the lesion evidence mentioned earlier (Johnsrude et al., 2000), as it points to coarser and finer resolution in left and right auditory cortex, respectively.

In addition to these fMRI results, a number of studies using MEG have provided evidence concerning specialization for pitch processing. Because of the increased temporal resolution afforded by MEG, these studies have been able to isolate two separate components: a transient "pitch-onset response" (Krumbholz et al., 2003) and a sustained component (Gutschalk et al., 2002; Gutschalk et al., 2004) of the auditory evoked potential that indicate pitch processing. Because these potentials have longer latencies than typical responses to sound onset, they have been interpreted as indicating a hierarchical process, in which pitch extraction follows initial sound onset–related processing, or in which the computation time is longer for pitch extraction. In agreement with the fMRI and lesion data mentioned previously, when the sources of these components are modeled, they are most frequently found to lie within lateral HG. Additional relevant electrophysiological data have also recently been provided by a case study of an epilepsy patient in whom an electrode was placed within HG (Schönwiesner & Zatorre, 2008); a pitch-onset response was

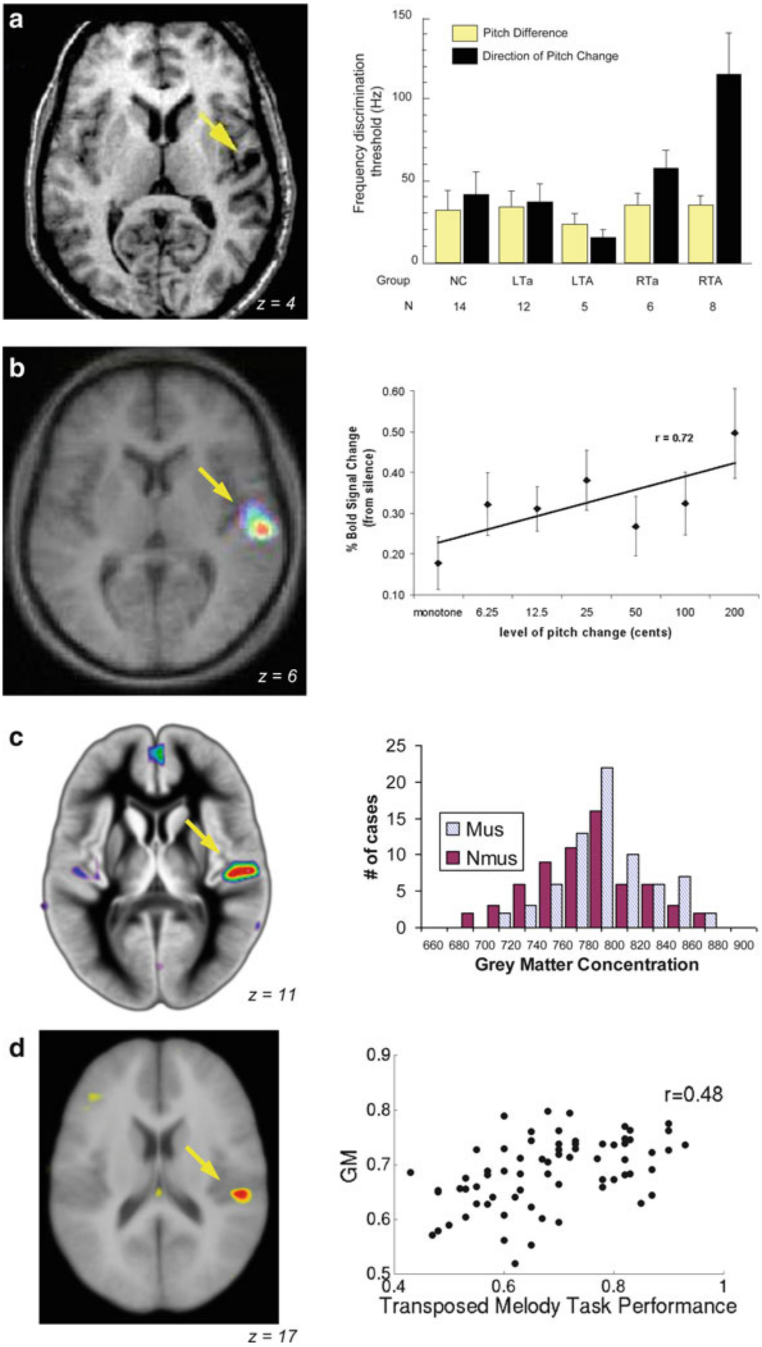


Fig. 10.2 Correspondence across studies of area in lateral portion of right Heschl's gyrus (HG; yellow arrows) in association with pitch-related functions. (**a**) Lesion study showing significantly enlarged discrimination thresholds for pitch direction judgments (black bars) in patients whose

elicited in the lateral portion of HG when a stimulus contained pitch, whereas the medial portion of HG instead showed a sound-onset response but no change when periodicity was introduced. This double dissociation thus supports the earlier findings indicating two separate responses, and confirms the differential localization of these sources (cf. Griffiths et al., 2010 for electrophysiological evidence challenging the notion of a "pitch center"). A lateral HG location for a pitch-sensitive area is also reported by Chait et al. (2006) using dichotic pitch, a very different type of pitch-eliciting stimulus, further supporting the conjecture that pitch percepts of different origin are calculated or represented in a particular cortical field.

10.4 Neural Substrates of Melodic Processing

Identification of a neural region that displays pitch sensitivity merely represents the first step in understanding a very complex set of neural operations that can lead to something like recognition of a melody. Having extracted pitch information from a periodic signal, the relationships between these pitches must be computed. Cognitive psychology confirms and expands upon music theory concepts that state that the relationships between pitches (i.e., the frequency ratios, known as intervals) are critical in defining a melody, and not the absolute pitches of the tones comprising the melody (Attneave & Olson, 1971). More precisely, the initial encoding of a novel melody relies on two other features of tonal relationships: the pattern of rising and falling intervals, known as the contour, and the scale in which the tune is played (Dowling, 1978). These interval relationships are encoded in long-term memory and form the basis for our ability to recognize a tune regardless of the musical key in which it is performed, a process known as transposition.

The neural substrates for these aspects of melodic processing have been investigated both with lesion studies and functional imaging. Two consistent general findings have emerged from this literature: that areas along the superior temporal gyrus both anterior and posterior to HG are important for melodic processing, and that there is a relative asymmetry favoring structures on the right side for such processing. Several studies have examined the perception of melodic patterns in patients with superior temporal gyrus (STG) lesions anterior to HG, and have shown that these

Fig. 10.2 (continued) excisions included right HG (group labeled RTA) but not in patients with more anterior resections (RTa) or left-sided damage (LT). Thresholds for same/different pitch discrimination were unaffected (yellow bars). (Adapted from Johnsrude et al., 2000.) (**b**) fMRI data showing blood oxygenation signal increases as a function of parametrically increasing pitch change in a tonal pattern. (Adapted from Hyde et al., 2008.) (**c**) Structural MRI study showing voxel-based morphometry contrast of highly trained musicians versus nonmusicians; right panel shows distribution of gray-matter concentration scores at the site of maximal difference. (Adapted from Bermudez et al., 2009.) (**d**) Structural MRI study showing relationship between task performance on a melodic discrimination task and gray matter concentration values at peak location. (Adapted from Foster & Zatorre, 2010.)

lead to specific impairments in perceptual processing of melodies (Zatorre, 1985; Liégeois-Chauvel et al., 1998; Stewart et al., 2006); the effects are typically bilateral but with greater deficits observed after damage on the right side.

Early functional neuroimaging studies of melody perception showed that cortical areas in the STG outside HG are active, depending on the control condition used (Zatorre et al., 1994; Griffiths et al., 1999; Binder et al., 2000). A common finding in these studies is that areas both anterior and posterior to HG are recruited (Griffiths & Warren, 2002; Zatorre et al., 2002a). The regions posterior to HG, usually including the planum temporale (PT), are sensitive to frequency modulation in general, not only in the context of melodies (Thivard et al., 2000; Hall et al., 2002; Hart et al., 2003). The posterior auditory cortex is also more sensitive to pitch height, which refers to the spectral weighting of a sound; conversely, more anterior areas show sensitivity to pitch chroma, which is a feature related to the relative position of a pitch within a scale (Shepard, 1982; Krumhansl, 1990; Warren et al., 2003).

The relationship of these regions anterior and posterior from HG to the pitch-sensitive area described above remains to be determined, but it seems likely that both anterior STG and the PT receive input from it. In keeping with the idea of hierarchical processing derived from the anatomy, we may assume that these more downstream regions perform computations beyond pitch extraction, perhaps related to scale, interval size, and contour, which, as noted previously, are crucial for perception. One study (Patterson et al., 2002) is directly relevant here because it dissociated neural activity originating from the lateral portion of HG (during processing of simple pitch) from activity in posterior and anterior STG areas, which was more sensitive to processing of melodies, consistent with the proposal that these distal regions are involved in higher-order feature analysis of melodic information. Another study that gives a significant insight into the role of posterior auditory cortex for processing melodic patterns used a parametric approach to vary the amount of informational content, or entropy, in a sequence of tones (Overath et al., 2007). The principal finding was that the PT was most sensitive to the transitional probability of the tone patterns, such that "… sequences are encoded by a mechanism that demands less computational resource for … low information content and high redundancy (due to the predictability of the sequence) than that required to encode sequences with little or no redundancy" (p. 2728). This conclusion is also in line with the general concept of hierarchical processing, as mentioned previously.

It is clear from the literature, then, that melodies are not processed in a simple sequential manner, but rather that they involve setting up complex cognitive schemas. A good example of what this means comes from considering the expectancies generated by tonal music, which embody listeners' implicit knowledge about musical rules acquired from listening to music in their culture, as well as their explicit knowledge about particular musical events based on prior exposure to specific pieces of music (Huron, 2006). Based on such types of prior knowledge, hearing a certain set of notes leads one to expect certain other notes, which is why even nonmusicians can normally detect inappropriate notes in tonal music, even if it is novel.

The reason this phenomenon is of interest is because it points to the ability to predict upcoming events based on past regularities, which is an obviously adaptive capacity for the nervous system. Music provides an excellent way to probe the cortical substrates associated with such predictive mechanisms, and one way to do so is to look for markers of violations of musical expectancies. This has been done extensively in the electrophysiological literature using the so-called mismatch negativity (MMN) response (review: Näätänen et al., 2007 and Alain and Winkler, Chapter 4), a signal associated with a stimulus that does not fit with some prior context, and that is relatively automatic (i.e., independent of attention). The MMN was first described in terms of simple situations, such as when a sound deviates from a repetitive background; however, it turns out that it is elicited in a wide array of circumstances, and of greatest interest from a cognitive musical perspective, appears to be sensitive to higher-order regularities. For example, some studies have shown that an MMN can be elicited to violations of contour, irrespective of the actual pitch values used (Tervaniemi et al., 2001); conversely, a change in the interval size of a melody without a contour change can also evoke a MMN (Trainor et al., 2002), indicating that both contour and interval size are encoded as part of the processing of a melody, in keeping with cognitive principles mentioned earlier (Dowling & Harwood, 1986). The mismatch-negativity approach has also been used to examine the processing of simultaneously presented musical events, which of course is commonplace in real music. Fujioka and colleagues (2005) presented listeners with two simultaneous melodies such that a deviation was present in either one or the other; this situation elicited an MMN showing that multiple separate representations of melodies can be held at one time.

The precise cortical localization of the MMN is not easy to ascertain, especially because the pattern of electrical activity associated with it changes depending on the feature that is varied (Giard et al., 1995); however, there is good evidence from fMRI that it does originate from several regions of auditory cortex, as well as frontal cortex (Opitz et al., 2002; Molholm et al., 2005; Rinne et al., 2005; Schönwiesner et al., 2007). These studies suggest that increasingly more abstract levels of change detection are encoded as one ascends from primary auditory cortex to surrounding auditory cortical regions, to frontal cortex, in keeping with the concept of hierarchical organization. Thus, responses to changes in low-level features of a sound can be elicited from medial HG (and indeed, have even been reported to occur subcortically, Kraus et al., 1994; Schönwiesner et al., 2007), whereas responses to more abstract features originate in more distal portions of the auditory processing stream, including regions of inferior or dorsolateral frontal cortex.

A good example of frontal-lobe involvement in detection of changes in abstract features is provided by studies of musical syntax, in which a series of chords follows a harmonic progression that is typical of Western tonal music. When an unexpected chord is introduced, which is itself consonant but is not a typical continuation of the progression, a negativity response occurs somewhat later than that coming from auditory cortex, has an anterior distribution, and is typically stronger on the right (Koelsch et al., 2000; Koelsch et al., 2003; Leino et al., 2007). MEG data localize the anterior negativity to inferior frontal cortex (Maess et al., 2001), a finding confirmed with fMRI data (Koelsch et al., 2002), although temporal areas are also likely involved (Tillmann et al., 2006). These findings can be interpreted as indicating that interactions between sensory processing regions in the hierarchically organized ventral stream of auditory processing (see later) interact with inferior frontal cortex to generate representations of structural regularities; this idea also has parallels in linguistic processing (Friederici et al., 2003; Patel, 2003).

10.5 Hemispheric Asymmetries of Auditory Processing

Human auditory cortex shows a relatively consistent morphological asymmetry. Several studies using structural neuroimaging techniques have documented that the left HG has a larger volume than the right (Penhune et al., 1996, 2003; Dorsaint-Pierre et al., 2006), a finding that is consistent with some of the earliest anatomical observations of this structure (von Economo & Horn, 1930). This phenomenon is most evident in the white matter underlying the gyrus, suggesting that it may be related to a greater number of fibers and/or increased myelination (Sigalovsky et al., 2006). Several human postmortem studies have lent further evidence favoring the existence of auditory cortex asymmetries in a number of microstructural features, including total volume and myelin thickness (Anderson et al., 1999), spacing of intrinsic connections (Galuske et al., 2000), microcolumn organization (Seldon, 1981; Chance et al., 2006), and pyramidal cell size (Hutsler & Gazzaniga, 1996). Although it is unknown precisely how these anatomical features relate to the types of functional processing differences most relevant to our discussion, it is nonetheless likely that structure–function relationships are important (for some analysis, see Giraud and Poeppel, Chapter 9). One of the few studies to examine this question directly via functional magnetic resonance imaging (fMRI; Warrier et al., 2009) did show that larger volumes of left HG were associated with larger extents of cortex sensitive to temporal cues on the left, whereas larger volumes of right HG were associated with larger extents of spectral-related cortex on the right. This demonstration, together with other structure–behavioral correlations discussed in greater detail later (Schneider et al., 2002; Golestani et al., 2007; Wong et al., 2008; Foster & Zatorre, 2010a) indicate that the anatomical organization of auditory cortex is important in understanding its functional role in music, speech, and other auditory processes.

Asymmetries of function favoring right auditory cortex for pitch and melodic processing have been observed since the earliest lesion studies (Milner, 1962), and also documented in many subsequent lesion studies as indicated previously. The neuroimaging literature is also generally in agreement that asymmetries associated with tonal processing beyond the level of HG and surrounding areas point to specialization favoring the right side, and rarely in the opposite direction; this finding has been confirmed for a wide variety of paradigms and imaging modalities (e.g., Zatorre et al., 1994; Tervaniemi et al., 2006; Herholz et al., 2008) that is not reviewed in detail here. However, there is less certainty about whether such asymmetries can also be observed at the earliest stage of cortical processing. For example, in Patterson et al. (2002), hemispheric differences favoring the right side emerged only in areas outside HG, and only for melodic stimuli and not simple pitch-bearing stimuli.

The results of several related studies have also shown symmetric responses in areas in or near HG (Griffiths et al., 1998; Penagos et al., 2004; Hall et al., 2006).

However, right-sided predominance has been consistently observed in core or adjacent belt areas under specific circumstances with a number of different stimuli, particularly those involving fine-grained spectral processing (e.g., Zatorre & Belin, 2001; Jamison et al., 2006); thus the presence of asymmetries early in the processing stream may depend critically on specific stimulus and/or task parameters. For example, Hyde et al. (2008) reported that asymmetries of hemodynamic responses emerged only when pitch variation in a tonal pattern was relatively small; in contrast, when the stimulus contained larger pitch changes (on the order of 200 cents, or two semitones) the response was symmetrical and bilateral. Similarly, Schönwiesner et al. (2005; see also Warrier et al., 2009) demonstrated that even with nonperiodic (noise) stimuli, asymmetries favoring the right side were elicited most clearly in terms of sensitivity to bandwidth, whereas responses from left auditory cortex emerged more as a function of temporal factors, again suggesting a specialization for fine-grained spectral processing.

A critical point to make is that in essentially all of these studies asymmetries are of a relative, and not absolute nature. Thus, we may conclude that even at the level of core and adjacent belt cortices there do exist functional asymmetries, but that they tend to be subtle, and that they emerge only if the experimental conditions are set up to be sensitive to them. The theoretical import of these findings is that they point to a specialization at relatively early stages of processing, favoring the analysis of fine-grained spectral information in right auditory cortex; this in turn would play a critical role in processing musical information, because as indicated earlier, precise pitch relationships are essential to musical processing in a way that has no equivalent in the speech domain.

One model to explain the presence of these asymmetries posits that hemispheric differences arise from a fundamental constraint in the ability of auditory cortices in each hemisphere to optimize processing in the temporal as compared to the spectral domain (Zatorre et al., 2002a). According to this view, resolution of fine differences in the frequency domain are accompanied by a relatively poorer resolution in the temporal domain, because it takes longer to sample enough incoming information to achieve the higher spectral definition; conversely, rapid sampling of the signal is necessary to achieve high temporal resolution, but this comes at the expense of spectral resolution. The right and left auditory cortices would thus have complementary properties such that the right would be relatively more specialized for resolving small frequency differences, whereas the left would be relatively more specialized for resolving small temporal differences. A related model stipulates that the two auditory cortices differ in terms of temporal integration windows, with the left and right having shorter and longer time integration constants, respectively (Poeppel, 2003 and see Giraud and Poeppel, Chapter 9, for some discussion). To the extent that speech and music exploit different ends of this temporal–spectral continuum (speech is a broadband signal containing rapid transients; music typically contains precise frequency relationships but its elements change comparatively slowly), it would then follow that they demonstrate a tendency to show opposite hemispheric weightings, a conclusion in accord with the preponderance of the evidence.

Several pieces of empirical evidence support this type of model, although the paradigms are not necessarily directly related to the processing of musical stimuli. The imaging evidence from experiments in which spectral and temporal features of sounds are manipulated is consistent with these proposals (Zatorre & Belin, 2001; Schönwiesner et al., 2005; Overath et al., 2008). Further consistent data come from an MEG experiment in which the overall stimulus features were kept constant, but sensitivity to change in either temporal or spectral characteristics was found to engage preferentially left or right auditory cortex, respectively (Okamoto et al., 2009). Boemio and colleagues investigated the concept of temporal integration differences across auditory cortices in each hemisphere more directly (Boemio et al., 2005). These investigators parametrically varied the segment transition rates in a set of concatenated narrow-band noise stimuli, such that the durations varied from quickly (12 ms) to slowly changing (300 ms). Sensitivity to this parameter was observed in primary and adjacent auditory cortices symmetrically on both sides, but the more slowly modulated signals preferentially recruited the right superior temporal sulcus (STS). The authors interpret this finding to mean that "…there exist two timescales…with the right hemisphere receiving afferents carrying information processed on the long timescale and the left hemisphere those resulting from processing on the short timescale" and they hence conclude that this is "… consistent with the proposal suggesting that left auditory cortex specializes in processing stimuli requiring enhanced temporal resolution, whereas right auditory cortex specializes in processing stimuli requiring higher frequency resolution" (p. 394).

Another related study that bears on this question used simultaneous electroencephalography (EEG) and fMRI to show that spontaneous EEG power variations at a relatively fast oscillatory rate correlate best with left auditory cortical activity, whereas fluctuations within a slower rate correlate best with activity in right auditory cortex (Giraud et al., 2007; see also Morillon et al., 2010). Giraud and colleagues interpret these findings as indicating that this differential functionality could underpin left hemispheric speech processing because of its reliance on rapid temporal features, whereas integration over slower time windows (on the order of 100–300 ms) in right auditory cortex "…optimizes extraction of slow information, thus promoting processing of periodicity…and music" (p. 1130). Thus, these findings suggest that asymmetries at early stages of auditory processing may be related to intrinsic properties of these cortical circuits within each hemisphere, adding further evidence in favor of the idea that hemispheric differences may be conceptualized in terms of differences in their capacity to resolve spectral versus temporal features. It should also be appreciated that the asymmetric response can be influenced by, or even be entirely dependent upon task demands and other contextual factors. A good example of this effect is provided by Brechmann and Scheich (2005), who presented listeners with frequency–modulated tones of different durations, and asked them to make judgments either of pitch direction or of duration. Categorization of pitch direction increased hemodynamic activity in right posterior auditory cortex, whereas duration categorization increased activity in left posterior auditory cortex, indicating that the hemispheric asymmetries predicted by the models described previously can sometimes emerge only under certain task states.

10.6 Dorsal-Stream Model of Auditory Processing and Its Relation to Music: Where, How, or Do?

As mentioned earlier, auditory regions both anterior and posterior to HG are more sensitive to frequency modulations and in particular, higher-order pitch features such as pitch height, pitch chroma, and pitch contours in melodies. Distinct patterns of anatomical connections within auditory cortex suggest that there are at least two auditory processing streams stemming from primary auditory cortex, each of which may contribute to processing different higher-order aspects of auditory stimuli. In monkeys, the more anterior areas of primary auditory cortex send the densest projections to anterior portions of secondary, tertiary, and other higher-order areas of auditory cortex, while the posterior region of primary auditory cortex projects mainly to posterior regions of higher-order auditory cortex; there are relatively few connections between anterior and posterior areas (Hackett et al., 1998; Kaas & Hackett, 2000). These divergent projections also target different subdivisions of extratemporal regions—anterior auditory areas project to more anterior and orbitofrontal regions of prefrontal cortex, while posterior auditory cortex is interconnected with more posterior areas of STG and STS, posterior parietal cortex, premotor cortex, and dorsolateral frontal cortex (Cavada & Goldman-Rakic, 1989; Hackett et al., 1999; Romanski et al., 1999).

These distinct pathways observed within auditory cortex and electrophysiological studies of audition in primates (e.g., Recanzone et al., 2000; Tian et al., 2001) gave rise to a dual-stream model of auditory processing (for more details, see Rauschecker & Tian, 2000). Similar to the proposed dual-stream model for the visual system (Ungerleider & Mishkin, 1982; Ungerleider & Haxby, 1994), the (postero)dorsal stream was originally suggested to mediate auditory spatial processing (the “where” pathway), while the (antero)ventral stream was associated with processing auditory object information (the “what” pathway), including features of species-specific vocalizations (Rauschecker & Tian, 2000).

Considerable lesion and functional neuroimaging evidence supports the proposed ventral/dorsal distinction (e.g., Alain et al., 2001; Clarke et al., 2002; Warren & Griffiths, 2003), especially for the ventral stream (Zatorre et al., 2004), but the role of the dorsal stream in spatial processing is more controversial because it is not clear that it is specifically sensitive to spatial location (Zatorre et al., 2002b; Smith et al., 2010). Belin and Zatorre (2000) suggested that to make a better analogy with vision, the dorsal stream should be characterized as being involved in processing changes in spectral energy over time, which makes this stream more relevant for melodic contour processing (the “where in frequency” or “how” pathway). Romanski et al. (2000) argued that the dorsal pathway processes both spatial elements and spectral changes to detect auditory motion. To reconcile these incongruent views, Warren and colleagues (2005) proposed that the dorsal auditory stream serves an auditory–motor function (the “do” pathway), which resembles Goodale and Milner’s (1992) sensorimotor model for the dorsal visual stream in which visual information is used to guide motor output. In Warren’s model, auditory spatial information can be extracted to prepare for movement either toward or away from the auditory stimulus, and alternatively, externally presented speech is processed and matched with the proper response template; these "auditory–motor transformations" can then be sent via a dorsal auditory pathway to prefrontal and premotor cortex. Hickok and Poeppel (2000, 2004, 2007) also put forth an auditory–motor model of speech processing in which the dorsal stream transforms auditory representations of speech into motor programs for speech (see Hickok and Saberi, Chapter 12); however, unlike Warren's model, the authors indicated that their model is not amenable to processing auditory spatial information. More recently, Rauschecker and Scott (2009) have adapted the original what/where model of auditory processing to include the growing body of evidence that implicates the dorsal stream in speech processing and control. This revised model posits that speech is in part processed via the ventral stream and converted to motor representations in the inferior frontal gyrus and ventral premotor cortex, which varies from both the Warren et al. and Hickok/Poeppel models that attribute auditory–motor transformations to the dorsal stream alone. However, in this new Rauschecker/Scott model, an efference copy of the motor program is then sent from the premotor areas to inferior parietal cortex in the dorsal auditory stream for comparison of the predicted outcome and actual feedback stemming from auditory cortex. In addition, Rauschecker and Scott suggest that inferior parietal cortex, which has also been associated with attention or intention, can influence the selection of motor programs in premotor areas. In light of these proposed models, recent neuroimaging studies in music cognition provide evidence that higher-order processing of musical stimuli, apart from pitch and melodic contour extraction, engages regions within the dorsal auditory stream; these studies help to forge a consensus across the literature and to integrate the various proposed models.

Returning to the shopping-mall jingle presented at the beginning of this chapter, one may ask how a listener so readily recognizes a tune when it is presented in a different key (i.e., different pitch level) than when it was first encoded. Recent studies have suggested that this process engages the intraparietal sulcus (IPS), which forms part of the dorsal stream discussed earlier. The IPS is recruited when listeners make judgments about pairs of melodies that have been transposed, as compared to melodies in the same key; furthermore, the degree of activation correlates with behavioral performance on the task, indicating a fairly direct relationship between brain activity and ability to transpose (Foster & Zatorre, 2010b). One interpretation of this finding is that the IPS is recruited because it supports a general capacity for transformation of systematically related stimulus attributes from one frame of reference to another. Converging evidence that the dorsal stream is involved in musical transformation comes from a study showing IPS activation when listeners made judgments of tunes that had been temporally reversed (Zatorre et al., 2010). The IPS activation, which overlapped with that of the transposition study, in this case would be related to transformation in a temporal rather than a pitch-based coordinate frame.

The concept that the dorsal stream is important for manipulation and transformation of musical inputs fits in well with the models proposed in the preceding text, in which the dorsal stream is involved with spatial localization and control of action, because these mechanisms are thought to underlie the transformation of information across reference frames. In the visual domain there is much evidence that the IPS is important for this type of processing, including such tasks as reaching/grasping and mental rotation (Grefkes & Fink, 2005; Culham et al., 2006; Zacks, 2008). Additional evidence for a link between visual and auditory transformation tasks comes from behavioral findings of an association between behavioral ability in visual rotation and melody reversal (Cupchik et al., 2001); conversely, persons with congenital amusia (see Section 10.8 for more discussion) also show impaired visual mental rotation (Douglas & Bilkey, 2007; cf. Tillmann et al., 2010 for a failure to replicate). Because the IPS is a multisensory region that receives inputs from many cortical regions, including from posterior auditory cortices (Schroeder & Foxe, 2002; Frey et al., 2008), it is logical to suppose that it may carry out similar computations on different inputs.

If the dorsal stream is part of an action-oriented network that integrates multimodal feedforward and feedback information (Rauschecker & Scott, 2009), one might also expect it to be involved in auditory-guided motor acts. A good test of this idea is provided by examining vocal corrections for pitch production during singing. In an fMRI study of singing (Zarate & Zatorre, 2008), auditory feedback was pitch-shifted at specific times to mimic an incorrectly produced note, and participants were instructed to alter their singing pitch to fully correct for the shifted feedback; this particular task was designed to engage brain regions involved in voluntary vocal pitch regulation. Whereas singing with normal, unperturbed auditory feedback did not engage dorsal regions, the investigators found that the IPS was recruited only when pitch-shifts were presented. As illustrated previously with melodic transposition and reversal, the IPS may be engaged as participants evaluated the size and direction of the pitch-shift before voluntarily changing their singing pitch in response to the shifted feedback (Fig. 10.3A, left; Zarate & Zatorre, 2008). Indeed, singers showed enhanced functional connectivity between posterior auditory cortex and IPS as they voluntarily corrected for a large, 200-cent pitch-shift in auditory feedback, while smaller pitch-shifts engaged a more automatic neural system for vocal control (Fig. 10.3A, right; Zarate et al., 2010b). Overall, it may be argued that the dorsal pathway of auditory processing plays a role in calculating and comparing "how" pitch changes over time or relative to another tonal reference point before making a musically relevant decision (e.g., recognizing a pattern) or a specific motor response to these pitch transformations.

The information about musically relevant transformations stemming from higher-order auditory regions, including posterior auditory cortex and IPS, can be used by premotor cortex to prepare or "do" the correct motor program in response to the auditory stimulus (see Beurze et al., 2007; Mars et al., 2007; Chen et al., 2009). In general, the dorsal portion of premotor cortex (dPMC) has been implicated in associating auditory information with motor responses, especially within a musical context (review: Zatorre et al., 2007). In rhythm-tapping tasks, activity within the dPMC increases as rhythms become more salient and complex, and the functional coupling between dPMC and posterior auditory cortex increases as the metrical salience increases (Chen et al., 2006, 2008b). Together, these observations suggest that this

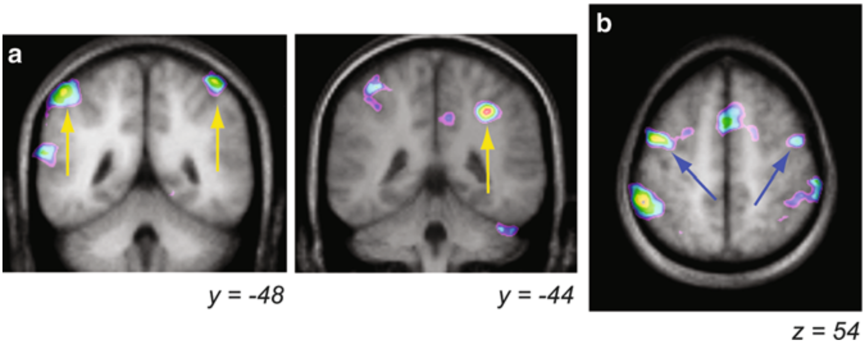


Fig. 10.3 Brain regions recruited during motor control of pitch. (**a**) Experienced singers and nonmusicians commonly recruited the intraparietal sulcus (IPS; yellow arrows) only during singing with pitch-shifted auditory feedback, presumably as pitch-shift direction and magnitude were evaluated. (Adapted from Zarate & Zatorre [2008; left] and Zarate et al. [2010b; right]. (**b**) Dorsal premotor cortex (dPMC; blue arrows), active in both singers and nonmusicians during singing with pitch-shifted feedback, may associate pitch-shift information with motor adjustments of pitch. (Adapted from Zarate & Zatorre, 2008.)

region may be involved in forming a conditional association between auditory cues (i.e., the rhythm) with highly organized motor responses (e.g., synchronized tapping; Chen et al., 2006, 2008b). Other subdivisions of premotor cortex also respond to rhythmic stimuli in different manners—in an fMRI study in which rhythms were first passively presented with no intention for reproduction via tapping, and then later sessions presented rhythms specifically designated for tapping, the mid-premotor cortex (mid-PMC) was recruited during both passive listening to rhythms and listening to rhythms with intention to tap, whereas the ventral premotor cortex (vPMC) was engaged only when presented rhythms are designated for tapping (Chen et al., 2008a). These authors proposed that mid-PMC "may have a more general role in attending to features of the physical stimulus, tracking the sequentially presented auditory events in the anticipation that they might be of relevance to the motor system" (p. 2853), and the selective recruitment of vPMC during "active" listening of rhythms targeted for tapping suggested that this region "maps a specific sound with a precise movement that produces that sound" (p. 2850).

Pitch-related musical tasks such as singing also engage different regions within the premotor cortex. As participants reproduce a target note via singing, the vPMC is active—perhaps due to the direct sound-to-action mapping—along with other brain regions within the functional network for singing (Perry et al., 1999; Brown et al., 2004b; Kleber et al., 2007; Zarate & Zatorre, 2008). However, once auditory feedback is pitch-shifted during singing, both the IPS and dPMC are engaged, the IPS may be recruited to assess the magnitude and direction of the pitch-shift (Fig. 10.3A), while the dPMC may associate this pitch-shift cue with specific vocal motor adjustments in response to the shifted feedback (Fig. 10.3B; Zarate & Zatorre, 2008).

To summarize, the auditory processing of music occurs in a hierarchical fashion, and each step may be attributed to specific brain regions within the dorsal auditory stream emanating from HG: pitch extraction within the right lateral HG, posterior auditory cortex processing melodic contour and rhythmic structure, and calculation and comparison of pitch relationships or temporal manipulation within IPS, before selecting a motor response to the musical stimuli within the premotor cortex. The studies cited here demonstrate that the sensitivity of IPS to calculation and comparison of "how" frequencies change over time supports the "how" or "where-in-frequency" model suggested by Belin and Zatorre (2000), whereas the use of auditory information to prepare motor responses to musical stimuli, which is attributed to the premotor cortex, lends credence to the auditory–motor "do" function of the proposed dorsal-stream models of Warren et al. and Hickok and Poeppel. Thus based on these observations, the dorsal auditory stream may lend itself toward a "how-to-do" function in auditory processing.

10.7 Role of Training and Experience on Auditory Cortical Function and Structure

The research findings described begin to provide an outline of the processing pathways important for musical perception and production. This description would be incomplete, however, if one were to assume that these functional networks are static, and do not change as a function of various types of experience. Indeed, several decades of neuroscience research have established that short- or long-term training is associated with a variety of functional adaptations. For example, learning-induced improvements in perception are often accompanied by changes in cortical organization, such that experience with specific stimuli leads to an enhanced or expanded representation in the corresponding sensory cortex. This pattern has been consistently reported in neurophysiological studies of auditory learning in animals (e.g., Buonomano & Merzenich, 1998; Irvine, 2007; cf. Brown et al., 2004a for evidence that this is not necessarily always the case). Similar findings have emerged in human studies of auditory learning, although it is not at all clear that the phenomena being measured are actually directly comparable to what is measured in animal neurophysiological studies. Despite this uncertainty, which makes comparison across species and paradigms difficult, the human studies do converge in showing clear evidence of changes in auditory cortical responses as a function of experience.

10.7.1 Training Effects on Auditory Cortical Activity

Among the clearest such findings are studies in which training produces a greater amplitude of certain electrical or magnetic evoked potentials that are thought to originate from auditory cortex, such as, for example, the MMN, which is enhanced with both speech (Kraus et al., 1995) or musical training (Lappe et al., 2008).

This latter study is of particular interest here, as it showed not only an effect of training under controlled conditions (piano training over 2 weeks), but it also demonstrated that sensorimotor training (when subjects were trained to play and listen to a simple melody) was much more effective than a purely auditory training regime in eliciting the enhancements in the evoked response. This finding speaks to the importance of the sensorimotor interactions described in the preceding section. As well, this study showed that the enhancement was much larger over right than left auditory cortex, in keeping with much evidence previously discussed for specialization of right auditory cortex for pitch processing. Auditory training alone can, however, be effective in changing auditory cortical responses, as shown by Bosnyak et al. (2004), who trained adults to discriminate small frequency changes in amplitude-modulated pure tones, and found increases in several evoked auditory cortical responses, some of which were again most prominently observed in the right hemisphere (see also Menning et al., 2000 for related findings). More naturalistic training, in the form of music lessons, has also been shown to lead to significant enhancement of several auditory evoked potentials in children tested over a 1-year period (Fujioka et al., 2006).

Evidence of experience-dependent changes in auditory cortex activity has also been reported in several functional imaging studies. Studies in the speech domain showed auditory cortical enhancement to the same stimulus comparing activity before to after short-term exposure (Dehaene-Lambertz et al., 2005), or explicit short-term training (Golestani & Zatorre, 2004; Möttönen et al., 2006), with a predominance on the left side. Similar training-based studies in the tonal domain are few, and the results are less clear cut. Thus, whereas one study reported increased hemodynamic response in several brain areas, including auditory cortices, after training on a pitch-memory task (Gaab et al., 2006), another reported decreased auditory cortical activity after training on pitch discrimination (Jäncke et al., 2001). It is likely that both increases and decreases can be present, and that they reflect different task-related components, and/or different phases of learning (Kelly & Garavan, 2005).

10.7.2 Assessing Cortical Function in Musicians and Nonmusicians

Although controlled auditory-training studies using fMRI are not numerous, the findings in these specific training paradigms reviewed in the preceding text mirror the observations made in many studies contrasting musicians and nonmusicians, which provide another way to address the question of plasticity. The assumption here is that differences seen across groups reflect their different training histories, and thus changes may be interpreted in the context of experience-dependent plasticity. Although there is good evidence that this concept holds, it may not explain all of the phenomena observed for the simple reason that musical training in such studies is not assigned randomly to equally proficient groups; rather, people seek out musical training for a number of reasons, no doubt including certain proclivities and predispositions.

Among the first such observations between musicians and nonmusicians, Pantev and colleagues (1998) showed that the magnitude of the evoked magnetic response to a piano tone was greater among trained musicians than those without formal training; of particular relevance was the fact that a similar effect was not obtained for pure tones, suggesting that it was experience with piano sounds that led to the enhancement (because pure tones are not experienced outside of a laboratory). This finding argues for an experience-dependent plasticity, which is also supported by the observation that the size of the enhancement in the magnetic response correlates with age of commencement of training, such that those who began musical training earlier had a greater effect (Pantev et al., 1998). Further evidence for experience-dependent effects comes from a follow-up study in which it was observed that a similar enhancement could be specific to the type of musical instrument on which a musician received training (Pantev et al., 2001). Thus, there was a relatively greater response to violin than to trumpet tones in violinists, but the reverse pattern was found in trumpeters. Another example of enhanced neuromagnetic responses to tones in auditory cortex in a subsequent study (Schneider et al., 2002) was also linked to behavioral performance on a melodic task, suggesting that the cortical signal is relevant to musical ability. In a different domain altogether, Münte and colleagues (2001) showed that spatial attention in orchestra conductors, but not pianists or nonmusicians, was enhanced in peripheral auditory space, which was accompanied by a change in an electrical evoked potential that signals attentional deployment in the periphery. This special acuity in the auditory periphery presumably aids conductors in fine-tuning the orchestra's performance, especially if a particular musician and/or section deviates from the orchestral score.

The preponderance of the fMRI literature reports that there are greater hemodynamic changes in auditory cortex of musicians than nonmusicians when presented with certain (musical) stimuli or tasks. For example, comparing passive music listening to silence yields greater hemodynamic response in auditory cortices of musicians than nonmusicians (Ohnishi et al., 2001); however, the interpretation of such a finding is limited because without either a task or a stimulus control, it is difficult to know whether to attribute the effect to a general or specific change, or if it is an epiphenomenon due, for instance, to attentional changes, which are known to influence auditory cortical response (Hillyard et al., 1973; Petkov et al., 2004; Johnson & Zatorre, 2005).

A more targeted approach to investigating experience-dependent plasticity is to have musicians or nonmusicians perform a specific, controlled task such as judging whether a chord constitutes a good completion to a sequence or not (Koelsch et al., 2005); this manipulation yields evidence that anterior portions of the STG (along with frontal areas) are more active in both adults and children with musical training. This result allows one to conclude that a more well-developed processing of musical syntax exists among musicians, who are therefore more sensitive to violations of regularity in chord sequences, and that anterior auditory cortex plays a role in this process. Another approach that allows specific conclusions to be drawn about training-specific changes in cortical function is to compare the cortical responses to different stimuli. Thus, contrasting the hemodynamic response to flute versus violin music among musicians with training in each instrument leads to an enhanced response in auditory cortex as a function of such training (Margulis et al., 2009). This finding, similar to the MEG study mentioned earlier (Pantev et al., 2001), provides evidence that training can have rather specific effects on auditory cortical function. The precise mechanisms behind these changes, and what they may mean for the cognitive processes involved, remain to be fully understood, however.

Musical training also appears to have consequences outside of cortical areas. Several studies show that musicians have higher-amplitude brainstem evoked potentials to tones or periodic portions of speech sounds, and that they occur earlier (as early as 10 ms after a tone onset). The brainstem frequency following response, which entrains to stimulus periodicity and likely originates from the inferior colliculus, is also enhanced by musical training (Musacchia et al., 2007; Wong et al., 2007; Bidelman & Krishnan, 2009). These findings raise the intriguing possibility that cortical changes described earlier may also be related to changes occurring at the earliest input stages to the auditory system; alternatively, there may be interactions between cortical and subcortical mechanisms in terms of training-induced changes.

In the context of music production, several studies have also documented experience-dependent differences in auditory and motor cortical activity. When nonmusicians underwent short-term training to map particular pitches to certain keys to play short piano melodies, enhanced activity was observed after training in several regions, notably in auditory and motor cortices (Bangert & Altenmüller, 2003; Lahav et al., 2007). In an event-related potential study where auditory feedback was occasionally altered during keyboard performance of melodies, trained pianists showed a negative evoked potential in response to the altered feedback, whereas nonmusicians did not (Katahira et al., 2008). The investigators concluded that this negative potential reflected the mismatch between the intended auditory template created via feedforward mechanisms and altered feedback. Although nonmusicians displayed a late positive component in response to perturbed feedback (indicating detection of the feedback alteration), the absence of the negative potential reflected the lack of a feedforward auditory template upon which error detection is based, a phenomenon that the authors proposed may develop after training.

Several studies have examined training-related effects by testing experienced singers, who provide a unique opportunity to study auditory–vocal interactions. In a previously cited fMRI study that delivered pitch-shifted auditory feedback during singing tasks, both nonmusicians and singers recruited IPS and dPMC (within the dorsal auditory stream) as they voluntarily changed their vocal pitch to correct for the shifted feedback (Zarate & Zatorre, 2008). However, contrasting the hemodynamic responses between groups determined that nonmusicians displayed more activity within the dPMC than singers as they performed this vocal correction task, whereas singers recruited posterior auditory cortex, anterior cingulate cortex, and the anterior insula—different components of the dorsal auditory stream—more than nonmusicians. The authors argued that because both groups recruited the dPMC for this vocal correction task, this premotor region may serve as a basic interface for auditory–motor interaction. With training and practice, the experience-dependent network of posterior auditory cortex, anterior cingulate cortex, and insula—which are functionally connected to each other in singers and nonmusicians (Zarate & Zatorre, 2008; Zarate et al., 2010a, b)—may be engaged increasingly during vocal pitch regulation, as seen in experienced singers (Zarate & Zatorre, 2008). Yet, an fMRI investigation of short-term auditory training effects on vocal accuracy in nonmusicians reported that although short-term auditory training alone did improve perception, it was not sufficient to improve vocal accuracy or to engage these regions during singing tasks (Zarate et al., 2010a). The investigators suggested that both auditory and vocal motor training may be necessary to recruit the experience-dependent network observed in experienced singers, which resonates with previously mentioned MEG results that demonstrate cortical plasticity only after auditory–motor training, and not with auditory training alone (Lappe et al., 2008). Finally, the amount of musical experience may also contribute to enhancements in cortical activity during singing tasks—not only did opera singers display more activity within sensorimotor cortex compared to vocal students and amateur singers, but activity within the laryngeal and mouth representation of the somatosensory cortex also increased as a function of amount of singing practice (Kleber et al., 2010). This increase in somatosensory cortical activity, which the investigators suggest reflects better kinesthetic control of the vocal apparatus, coupled with enhanced processing of auditory feedback observed in Zarate and Zatorre's (2008) study, may help experienced singers perform singing tasks better than nonmusicians.

10.7.3 Effects of Musical Training on Neuroanatomy

The consequences of training and experience, as shown by the studies just reviewed, are not confined to cortical function, but seem to extend to anatomy as well. Using structural MRI techniques such as volumetry or voxel-based morphometry, several studies have documented increases in gray-matter concentration in a number of regions, including motor-related structures (motor cortex, supplementary motor area, cerebellum), and auditory cortex as well as some frontal regions in musicians (e.g., Sluming et al., 2002; Gaser & Schlaug, 2003). Similarly, increases in white-matter concentration have been found in structures such as the corpus callosum (Schlaug et al., 1995) and the corticospinal tract (Bengtsson et al., 2004) in musicians; moreover, the degree of change correlated with the age of commencement of training in these studies. Although these studies do not all necessarily converge on the precise areas in question, there is reasonable agreement that auditory and motor-related structures are the most consistently altered, and that these changes are related to the functional enhancements described earlier in auditory cortical response (and likely also related to changes evoked from the cortical representation of the fingers; Elbert et al., 1995; Schneider et al., 2002). More recently, cortical thickness, which is arguably a more specific measure than that derived from voxel-based morphometry, has been measured in musicians, and the findings tend to confirm the observations of the other methods, with thicker cortex revealed in auditory and motor cortex as well as frontal cortex (Bermudez et al., 2009). Importantly, cortical thickness, in right auditory cortex and parietal cortex has been shown to be predictive of performance on a melodic transposition task, but not on speech or rhythmic tasks (Foster & Zatorre, 2010a), indicating that the morphological features of specific cortical regions are directly relevant for aspects of behavior known to be mediated by those same regions.

An issue raised already for the functional studies is that cross-sectional comparisons of one group to another cannot determine causality. Indeed, in the paper just cited showing that cortical thickness predicts behavioral performance (Foster & Zatorre, 2010a), the authors suggest that although years of training does explain some of the variance, training alone may not be sufficient to account for the observed correlations, because when amount of training is controlled, the effect remains, suggesting that other factors beyond training, including perhaps preexisting dispositions, may play a role. However, investigations in which subjects undergo training and are studied before and after help to address this point much more directly. One such study (Hyde et al., 2009) examined the effect of naturalistic musical training in a group of children over the course of 15 months of training, and demonstrated not only that there are changes in right auditory cortex and motor cortex, as expected from the literature on adult musicians, but importantly, that the degree of change was predictive of performance, such that changes in auditory or motor regions correlated with behavior on auditory or motor tasks, respectively. Related data from the same cohort of children demonstrated that children who were highly practiced in music had larger anterior corpus callosum volumes and better performance on a motor task, compared to children with less musical practice and controls (Schlaug et al., 2009). It should be kept in mind, however, that in these studies, as in related longitudinal studies using EEG (Fujioka et al., 2006), assignment to training group is not done randomly, and so preexisting factors may still be playing a role to the extent that children who take music lessons may well already have some skill, or interest in music that distinguishes them from the control sample.

Experience-dependent changes are also known to relate to the developmental time window during which training occurs. Evidence from both animals and humans indicates that there may be “sensitive” periods in development when specific training can contribute to long-lasting changes in behavior and brain function (Knudsen, 2004; Dahmen & King, 2007; Kral & Eggermont, 2007). Many of the studies mentioned above report that the functional and/or structural effects measured are greatest among those who receive early training. It is not yet clear whether the age of commencement of training interacts with the duration of training to produce some of these effects, or if they are additive. However, behavioral studies in which number of years of training is controlled have shown that musicians who begin training earlier in life have distinct advantages in various sensorimotor tasks (Watanabe et al., 2007; Bailey & Penhune, 2010), suggesting that whatever neural changes underlie these behavioral advantages are likely to be either greater in magnitude or different in nature than the changes measured after training in later childhood or adulthood.

The findings reviewed in this section make it clear that musical training does have specific effects on brain structure, a finding also in line with training studies in nonmusical domains (e.g., Draganski et al., 2004). Taken together with the specificity (e.g., timbre-related effects) outlined in the functional studies reviewed earlier, we may confidently conclude that these effects are markers of experience-dependent plasticity. However, there is no reason to suppose that such experience-driven effects preclude the existence of predisposing factors. Indeed, in the verbal domain, several studies have suggested that preexisting variation in auditory cortical structure can have predictive value with respect to learning potential for distinguishing foreign speech sounds (Golestani et al., 2007; Wong et al., 2008). And as already mentioned, amount of training does not seem to explain all of the variance in the relation between auditory cortical structure and melodic task performance (Foster & Zatorre, 2010a). The most likely scenario, therefore, is that anatomical predispositions may influence some aspects of the outcome of training, while training in turn modifies those very anatomical features, hence resulting in a recursive loop, the details of which no doubt will generate sufficient work to understand, that it will keep future investigators busy for some time.

10.8 Amusia

The previous section concentrated on the enhanced musical processing that is associated with training in music as a means of understanding its neural basis. It is equally valuable to examine musical processing disorders to gain insight into the organization of musically relevant neural processes. The study of acquired musical perception or production disorders after brain damage goes back to the beginnings of neuropsychology (Critchley & Henson, 1977), but often these early reports were not useful because of their unsystematic and anecdotal nature. More specific knowledge was gained from experimental studies of brain-damaged individuals (Stewart et al., 2006), as already mentioned.

The class of musical disorders termed tone-deafness, or more specifically congenital amusia, is of particular interest because it arises as a developmental disorder in the absence of any gross accompanying cognitive impairment (Ayotte et al., 2002; Peretz & Hyde, 2003; Foxton et al., 2004). It results in a fairly specific perceptual problem in processing of pitch, as compared to temporal cues (Hyde & Peretz, 2004). Moreover, amusics are largely free from any impairments in nonmusical auditory processing, including speech, with the possible exception of intonation contours (Patel, 2008). It is also thought to have a genetic component (Drayna et al., 2001; Peretz et al., 2007).

Recent neuroimaging data on this disorder provide some important converging evidence for some of the findings reported earlier concerning the neural substrates for musically relevant processes. An obvious hypothesis to be investigated, based on the literature, was that congenital amusia should be associated with some kind of disorder within auditory cortex, precluding correct encoding of pitch information. Initial indications were that evoked-potential responses to deviant tones measured from auditory cortex were indeed anomalous (Peretz et al., 2005), suggesting that the core deficit might be located in auditory regions. However, more recent research nuances this conclusion by indicating that sensitivity does exist in early components of the evoked response to fine pitch variation in amusics, but the later, more cognitive components are absent, suggesting that the core deficit might be at later processing stages, and not in initial auditory cortical processing (Peretz et al., 2009).

Recent structural imaging data clarify the possible sources of the auditory processing deficit. Voxel-based techniques first identified changes in the white matter underlying the right inferior frontal gyrus (Hyde et al., 2006), which would be in keeping with the important role for this region in musical processing networks, as reviewed earlier (see also Mandell et al., 2007). A subsequent study using cortical thickness as the measure (Hyde et al., 2007) confirmed that right inferior frontal cortex was involved, but indicated a thicker cortex in this region, and also in a right superior temporal cortical region; the investigators interpreted this finding as possible evidence for a cortical malformation, similar to that seen in migrational disorders associated with developmental disorders such as dyslexia (Galaburda et al., 1985). The identification of both frontal and temporal anomalies, coupled with the white matter findings, suggests that an important aspect of the disorder might be a disruption in connectivity. This idea has been tested directly using diffusion imaging techniques in a study (Loui et al., 2009) that reported that amusic individuals did indeed have a reduction in the number of fibers that interconnect auditory regions with frontal regions via the arcuate fasciculus, as measured using MRI-based tractography.

Finally, the idea of disconnection between frontal and temporal regions is also supported by functional imaging evidence: an fMRI study (Hyde et al., 2011) found that although the response to pitch variation within right auditory cortex was relatively normal in amusics, measures of functional connectivity failed to indicate a correlated hemodynamic response in right inferior frontal cortex, in contrast to controls in whom this connectivity was detected. The emerging conclusion from these studies taken together, then, is that although some abnormal early processing within temporal-lobe auditory cortex cannot be ruled out, the disorder more likely arises from an abnormal interaction in the right hemisphere between sensory-based processing in auditory cortex and higher-order processes that depend on inferior frontal regions. This circuit is likely involved in various aspects of tonal processing, as discussed earlier, and hence a disruption in how information is transferred across these regions might be expected to result in the perceptual problems typical of amusia.

Impaired tonal processing in congenital amusia may also lead to deficient tone production or singing. Although amusics generally sing less accurately in the pitch domain than matched controls, a few amusics still sing quite competently according to the rating criteria employed in several studies (Ayotte et al., 2002; Dalla Bella et al., 2009). Thus, although poor pitch production may be due to impaired pitch perception in congenital amusia, the presence of accurate pitch production in a few amusics may suggest a partial dissociation between these two domains. In accordance with this conclusion, recent studies demonstrated that amusics' pitch-production thresholds were significantly smaller than pitch-perception thresholds (Loui et al., 2008), which may contribute to their ability to sing pitch changes in the correct direction, even though their overall production of individual pitches was highly variable and their pitch-comparison performance was at chance (Hutchins et al., 2010).

Together, these observations in congenital amusia support a dual-stream auditory hypothesis for pitch perception and production. The hallmark of congenital amusia—impaired pitch perception—may be caused by a compromised ventral auditory pathway. On the other hand, because some amusics are better at pitch-change production than pitch perception, the dorsal stream, which may play a role in auditory–motor processes that underlie pitch production or singing as discussed earlier in this chapter, may be relatively spared in at least some persons with congenital amusia.

10.9 Summary

If we return to our original example of recognizing a familiar tune that has been rearranged into an advertising jingle, we may conclude that the studies presented in this chapter show that a complex set of neural processes is indeed necessary for this process to succeed. The lateral HG is associated with extracting individual pitches and pitch changes within the melody, whereas anterior STG and PT are involved in determining relationships between pitches to define the melody contour. Certain dorsal-stream structures (e.g., IPS and dPMC) can process this melodic information further to execute musically relevant tasks, such as comparing the advertising jingle's melody with the original song, tapping to the jingle's rhythm, or singing along with the melody. As reviewed in the preceding text, each of these processes may be lateralized to the right hemisphere, although this hemispheric specialization may be only relative in nature.

Studies in normal, musically trained, and congenital amusic volunteers provide insight into the proposed functional roles in the dual-stream auditory hypothesis. Impaired tonal processing in congenital amusia may be caused by abnormal connectivity between temporal and inferior frontal cortex within the ventral auditory stream; this reinforces the proposed object-processing ("what") role of this stream. As stated earlier, the dorsal stream has been implicated in pitch transformations for melody comparisons, beat-tapping, and vocal pitch control, which supports the putative "how" and "do" roles of the dorsal auditory stream—this stream can process how frequencies change over time before that information is used to perform a musical task. Moreover, the dorsal stream is less compromised in congenital amusia, which may account for some amusics' ability to sing accurately. Finally, musical training can significantly modulate the extent to which these pathways are engaged during music processing. Musical training enhances cortical activity within the ventral stream during assessment of musical-sequence violations, compared to nonmusicians (Koelsch et al., 2005). Short-term musical training results in enhanced coactivation of auditory and motor regions within the dorsal stream during a music production task (Bangert & Altenmüller, 2003; Lahav et al., 2007), whereas long-term auditory–vocal motor training recruits a different dorsal-stream network for vocal pitch regulation than that observed in nonmusicians (Zarate & Zatorre, 2008).

In many respects this chapter will serve more to highlight our lack of knowledge, or our imprecise models, than to demonstrate solid understanding of the neural mechanisms underlying musically relevant behaviors. We are far from achieving a coherent understanding of all of the subprocesses that lead to the seemingly effortless recognition of a simple tune. This chapter serves to illustrate the value of studying musical processes not only in their own right, but also because of the unique and often unexpected insights that they yield into the basic organizational principles of nervous system function.

References

Alain, C., Arnott, S. R., Hevenor, S., Graham, S., & Grady, C. L. (2001). "What" and "where" in the human auditory system. *Proceedings of the National Academy of Sciences of the USA*, 98(21), 12301–12306.

Anderson, B., Southern, B. D., & Powers, R. E. (1999). Anatomic asymmetries of the posterior superior temporal lobes: A postmortem study. *Neuropsychiatry, Neuropsychology, & Behavioral Neurology*, 12, 247–254.

Attneave, F., & Olson, R. K. (1971). Pitch as a medium: A new approach to psychophysical scaling. *American Journal of Psychology*, 84, 147–166.

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain*, 125(2), 238–251.

Bailey, J., & Penhune, V. (2010). Rhythm synchronization performance and auditory working memory in early- and late-trained musicians. *Experimental Brain Research*, 204, 91–101.

Bangert, M. W., & Altenmüller, E. O. (2003). Mapping perception to action in piano practice: A longitudinal DC-EEG study. *BMC Neuroscience*, 4(1), 26.

Belin, P., & Zatorre, R. J. (2000). 'What', 'where' and 'how' in auditory cortex. *Nature Neuroscience*, 3(10), 965–966.

Bendor, D., & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature*, 436(7054), 1161.

Bendor, D., & Wang, X. (2006). Cortical representations of pitch in monkeys and humans. *Current Opinion in Neurobiology*, 16(4), 391–399.

Bengtsson, S., Ehrsson, H., Forssberg, H., & Ullén, F. (2004). Dissociating brain regions controlling the temporal and ordinal structure of learned movement sequences. *European Journal of Neuroscience*, 19, 2591–2602.

Bermudez, P., Evans, A. C., Lerch, J. P., & Zatorre, R. J. (2009). Neuro-anatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. *Cerebral Cortex*, 19, 1583–1596.

Beurze, S. M., de Lange, F. P., Toni, I., & Medendorp, W. P. (2007). Integration of target and effector information in the human brain during reach planning. *Journal of Neurophysiology*, 97(1), 188–199.

Bidelman, G. M., & Krishnan, A. (2009). Neural correlates of consonance, dissonance, and the hierarchy of musical pitch in the human brainstem. *Journal of Neuroscience*, 29(42), 13165–13171.

Binder, J., Frost, J., Hammeke, T., Bellgowan, P., Springer, J., Kaufman, J., & Possing, J. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex*, 10, 512–528.

Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience*, 8(3), 389.

Bosnyak, D. J., Eaton, R. A., & Roberts, L. E. (2004). Distributed auditory cortical representations are modified when non-musicians are trained at pitch discrimination with 40 Hz amplitude modulated tones. *Cerebral Cortex*, 14(10), 1088–1099.

Brechmann, A., & Scheich, H. (2005). Hemispheric shifts of sound representation in auditory cortex with conceptual listening. *Cerebral Cortex*, 15(5), 578–587.

Brown, M., Irvine, D. R. F., & Park, V. N. (2004a). Perceptual learning on an auditory frequency discrimination task by cats: Association with changes in primary auditory cortex. *Cerebral Cortex*, 14(9), 952–965.

Brown, S., Martinez, M. J., Hodges, D. A., Fox, P. T., & Parsons, L. M. (2004b). The song system of the human brain. *Brain Research: Cognitive Brain Research*, 20(3), 363–375.

Buonomano, D., & Merzenich, M. (1998). Cortical plasticity: From synapses to maps. *Annual Review of Neuroscience*, 21, 149–186.

Cavada, C., & Goldman-Rakic, P. S. (1989). Posterior parietal cortex in rhesus monkey: II. Evidence for segregated corticocortical networks linking sensory and limbic areas with the frontal lobe. *Journal of Comparative Neurology*, 287(4), 422–445.

Chait, M., Poeppel, D., & Simon, J. Z. (2006). Neural response correlates of detection of monaurally and binaurally created pitches in humans. *Cerebral Cortex*, 16(6), 835–848.

Chance, S. A., Casanova, M. F., Switala, A. E., & Crow, T. J. (2006). Minicolumnar structure in Heschl's gyrus and planum temporale: Asymmetries in relation to sex and callosal fiber number. *Neuroscience*, 143(4), 1041–1050.

Chen, J. L., Zatorre, R. J., & Penhune, V. B. (2006). Interactions between auditory and dorsal premotor cortex during synchronization to musical rhythms. *NeuroImage*, 32(4), 1771–1781.

Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008a). Listening to musical rhythms recruits motor regions of the brain. *Cerebral Cortex*, 18(12), 2844–2854.

Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008b). Moving on time: Brain network for auditory-motor synchronization is modulated by rhythm complexity and musical training. *Journal of Cognitive Neuroscience*, 20(2), 226–239.

Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2009). The role of auditory and premotor cortex in sensorimotor transformations. *Annals of the New York Academy of Sciences*, 1169, 15–34.

Clarke, S., Bellmann Thiran, A., Maeder, P., Adriani, M., Vernet, O., Regli, L., et al. (2002). What and where in human audition: Selective deficits following focal hemispheric lesions. *Experimental Brain Research*, 147(1), 8–15.

Critchley, M., & Henson, R. A., Eds. (1977). *Music and the brain: Studies in the neurology of music*. London: Heinemann.

Culham, J. C., Cavina-Pratesi, C., & Singhal, A. (2006). The role of parietal cortex in visuomotor control: What have we learned from neuroimaging? *Neuropsychologia*, 44, 2668–2684.

Cupchik, G. C., Phillips, K., & Hill, D. S. (2001). Shared processes in spatial rotation and musical permutation. *Brain and Cognition*, 46(3), 373–382.

Dahmen, J. C., & King, A. J. (2007). Learning to hear: Plasticity of auditory cortical processing. *Current Opinion in Neurobiology*, 17(4), 456–464.

Dalla Bella, S., Giguere, J. F., & Peretz, I. (2009). Singing in congenital amusia. *Journal of the Acoustical Society of America*, 126(1), 414–424.

Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., & Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. *NeuroImage*, 24(1), 21–33.

Dorsaint-Pierre, R., Penhune, V. B., Watkins, K. E., Neelin, P., Lerch, J. P., Bouffard, M., & Zatorre, R. J. (2006). Asymmetries of the planum temporale and Heschl's gyrus: Relationship to language lateralization. *Brain*, 129(5), 1164–1176.

Douglas, K. M., & Bilkey, D. K. (2007). Amusia is associated with deficits in spatial processing. *Nature Neuroscience*, 10(7), 915–921.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, 85, 341–354.

Dowling, W. J., & Harwood, D. (1986). *Music cognition*. Orlando, FL: Academic Press.

Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., & May, A. (2004). Neuroplasticity: Changes in grey matter induced by training. *Nature*, 427, 311–312.

Drayna, D., Manichaikul, A., de Lange, M., Snieder, H., & Spector, T. (2001). Genetic correlates of musical pitch recognition in humans. *Science*, 291, 1969–1972.

Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B., & Taub, E. (1995). Increased cortical representation of the fingers of the left hand in string players. *Science*, 270, 305–307.

Foster, N. E. V., & Zatorre, R. J. (2010a). Cortical structure predicts success in performing musical transformation judgments. *NeuroImage*, 53(1), 26–36.

Foster, N. E. V., & Zatorre, R. J. (2010b). A role for the intraparietal sulcus in transforming musical pitch information. *Cerebral Cortex*, 20(6), 1350–1359.

Foxton, J. M., Dean, J. L., Gee, R., Peretz, I., & Griffiths, T. D. (2004). Characterization of deficits in pitch perception underlying 'tone deafness'. *Brain*, 127(4), 801–810.

Frey, S., Campbell, J. S. W., Pike, G. B., & Petrides, M. (2008). Dissociating the human language pathways with high angular resolution diffusion fiber tractography. *Journal of Neuroscience*, 28, 11435–11444.

Friederici, A. D., Ruschemeyer, S.-A., Hahne, A., & Fiebach, C. J. (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension: Localizing syntactic and semantic processes. *Cerebral Cortex*, 13(2), 170–177.

Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., & Pantev, C. (2005). Automatic encoding of polyphonic melodies in musicians and nonmusicians. *Journal of Cognitive Neuroscience*, 17, 1578–1592.

Fujioka, T., Ross, B., Kakigi, R., Pantev, C., & Trainor, L. J. (2006). One year of musical training affects development of auditory cortical-evoked fields in young children. *Brain*, 129(10), 2593–2608.

Gaab, N., Gaser, C., & Schlaug, G. (2006). Improvement-related functional plasticity following pitch memory training. *NeuroImage*, 31(1), 255–263.

Galaburda, A., & Sanides, F. (1980). Cytoarchitectonic organization of the human auditory cortex. *Journal of Comparative Neurology*, 190(3), 597–610.

Galaburda, A. M., Sherman, G. F., Rosen, G. D., Aboitiz, F., & Geschwind, N. (1985). Developmental dyslexia: Four consecutive patients with cortical anomalies. *Annals of Neurology*, 18(2), 222–233.

Galuske, R., Schlote, W., Bratzke, H., & Singer, W. (2000). Interhemispheric asymmetries of the modular structure in human temporal cortex. *Science*, 289, 1946–1949.

Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *Journal of Neuroscience*, 23(27), 9240–9245.

Giard, M. H., Lavikahen, J., Reinikainen, K., Perrin, F., Bertrand, O., Pernier, J., & Näätänen, R. (1995). Separate representation of stimulus frequency, intensity, and duration in auditory sensory memory: An event-related potential and dipole-model analysis. *Journal of Cognitive Neuroscience*, 7(2), 133–143.

Giraud, A. L., Kleinschmidt, A., Poeppel, D., Lund, T. E., Frackowiak, R. S., & Laufs, H. (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron*, 56(6), 1127–1134.

Golestani, N., & Zatorre, R. J. (2004). Learning new sounds of speech: Reallocation of neural substrates. *NeuroImage*, 21, 494–506.

Golestani, N., Molko, N., Dehaene, S., LeBihan, D., & Pallier, C. (2007). Brain structure predicts the learning of foreign speech sounds. *Cerebral Cortex*, 17(3), 575–582.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, 15(1), 20–25.

Grefkes, C., & Fink, G. R. (2005). The functional organization of the intraparietal sulcus in humans and monkeys. *Journal of Anatomy*, 207, 3–17.

Griffiths, T. D., & Warren, J. D. (2002). The planum temporale as a computational hub. *Trends in Neurosciences*, 25(7), 348–353.

Griffiths, T. D., Büchel, C., Frackowiak, R. S. J., & Patterson, R. D. (1998). Analysis of temporal structure in sound by the human brain. *Nature Neuroscience*, 1, 422–427.

Griffiths, T. D., Johnsrude, I. S., Dean, J. L., & Green, G. G. R. (1999). A common neural substrate for the analysis of pitch and duration pattern in segmented sound? *Neuroreport*, 10, 3825–3830.

Griffiths, T. D., Kumar, S., Sedley, W., Nourski, K. V., Kawasaki, H., Oya, H., et al. (2010). Direct recordings of pitch responses from human auditory cortex. *Current Biology*, 20(12), 1128–1132.

Gutschalk, A., Patterson, R. D., Rupp, A., Uppenkamp, S., & Scherg, M. (2002). Sustained magnetic fields reveal separate sites for sound level and temporal regularity in human auditory cortex. *NeuroImage*, 15(1), 207–216.

Gutschalk, A., Patterson, R. D., Scherg, M., Uppenkamp, S., & Rupp, A. (2004). Temporal dynamics of pitch in human auditory cortex. *NeuroImage*, 22(2), 755–766.

Hackett, T. A., Stepniewska, I., & Kaas, J. H. (1998). Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *Journal of Comparative Neurology*, 394(4), 475–495.

Hackett, T. A., Stepniewska, I., & Kaas, J. H. (1999). Prefrontal connections of the parabelt auditory cortex in macaque monkeys. *Brain Research*, 817(1–2), 45–58.

Hall, D. A., & Plack, C. J. (2009). Pitch processing sites in the human auditory brain. *Cerebral Cortex*, 19(3), 576–585.

Hall, D. A., Johnsrude, I. S., Haggard, M. P., Palmer, A. R., Akeroyd, M. A., & Summerfield, A. Q. (2002). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, 12(2), 140–149.

Hall, D. A., Edmondson-Jones, A. M., & Fridriksson, J. (2006). Periodicity and frequency coding in human auditory cortex. *European Journal of Neuroscience*, 24(12), 3601–3610.

Hart, H. C., Palmer, A. R., & Hall, D. A. (2003). Amplitude and frequency-modulated stimuli activate common regions of human auditory cortex. *Cerebral Cortex*, 13, 773–781.

Herholz, S. C., Lappe, C., Knief, A., & Pantev, C. (2008). Neural basis of music imagery and the effect of musical expertise. *European Journal of Neuroscience*, 28(11), 2352–2360.

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, 4(4), 131–138.

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, 92(1–2), 67–99.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews: Neuroscience*, 8(5), 393–402.

Hillyard, S., Hink, R., Schwent, V., & Picton, T. (1973). Electrical signs of selective attention in the human brain. *Science*, 182, 177–180.

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectaction*. Cambridge, MA: MIT Press.

Hutchins, S., Zarate, J. M., Zatorre, R. J., & Peretz, I. (2010). An acoustical study of vocal pitch matching in congenital amusia. *Journal of the Acoustical Society of America*, 127(1), 504–512.

Hutsler, J., & Gazzaniga, M. (1996). Acetylcholinesterase staining in human auditory and language cortices—regional variation of structural features. *Cerebral Cortex*, 6, 260–270.

Hyde, K. L., & Peretz, I. (2004). Brains that are out of tune but in time. *Psychological Science*, 15, 356–360.

Hyde, K. L., Zatorre, R. J., Griffiths, T. D., Lerch, J. P., & Peretz, I. (2006). Morphometry of the amusic brain: A two-site study. *Brain*, 129, 2562–2570.

Hyde, K. L., Lerch, J. P., Zatorre, R. J., Griffiths, T. D., Evans, A. C., & Peretz, I. (2007). Cortical thickness in congenital amusia: When less is better than more. *Journal of Neuroscience*, 27(47), 13028–13032.

Hyde, K. L., Peretz, I., & Zatorre, R. J. (2008). Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia*, 46(2), 632–639.

Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., & Schlaug, G. (2009). Musical training shapes structural brain development. *Journal of Neuroscience*, 29(10), 3019–3025.

Hyde, K. L., Zatorre, R. J., & Peretz, I. (2011). Functional MRI evidence of an abnormal neural network for pitch processing in congenital amusia. *Cerebral Cortex*, 21(2), 292–299.

Irvine, D. R. F. (2007). Auditory cortical plasticity: Does it provide evidence for cognitive processing in the auditory cortex? *Hearing Research*, 229, 158–170.

Jamison, H. L., Watkins, K. E., Bishop, D. V. M., & Matthews, P. M. (2006). Hemispheric specialization for processing auditory nonspeech stimuli. *Cerebral Cortex*, 16(9), 1266–1275.

Jäncke, L., Gaab, N., Wüstenberg, T., Scheich, H., & Heinze, H.-J. (2001). Short-term functional plasticity in the human auditory cortex: An fMRI study. *Cognitive Brain Research*, 12, 479–485.

Johnson, J. A., & Zatorre, R. J. (2005). Attention to simultaneous unrelated auditory and visual events: Behavioral and neural correlates. *Cerebral Cortex*, 15, 1609–1620.

Johnsrude, I. S., Penhune, V. B., & Zatorre, R. J. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain*, 123, 155–163.

Kaas, J. H., & Hackett, T. A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Sciences of the USA*, 97(22), 11793–11799.

Katahira, K., Abla, D., Masuda, S., & Okanoya, K. (2008). Feedback-based error monitoring processes during musical performance: An ERP study. *Neuroscience Research*, 61(1), 120–128.

Kelly, A. M. C., & Garavan, H. (2005). Human functional neuroimaging of brain changes associated with practice. *Cerebral Cortex*, 15, 1089–1102.

Kleber, B., Birbaumer, N., Veit, R., Trevorrow, T., & Lotze, M. (2007). Overt and imagined singing of an Italian aria. *NeuroImage*, 36(3), 889–900.

Kleber, B., Veit, R., Birbaumer, N., Gruzelier, J., & Lotze, M. (2010). The brain of opera singers: Experience-dependent changes in functional activation. *Cerebral Cortex*, 20(5), 1144–1152.

Knudsen, E. I. (2004). Sensitive periods in the development of the brain and behavior. *Journal of Cognitive Neuroscience*, 16(8), 1412–1425.

Koelsch, S., Gunter, T. C., & Friederici, A. D. (2000). Brain indices of music processing: "Nonmusicians" are musical. *Journal of Cognitive Neuroscience*, 13, 520–541.

Koelsch, S., Gunter, T. C., von Cramon, D. Y., Zysset, S., Lohmann, G., & Friederici, A. D. (2002). Bach speaks: A cortical "language-network" serves the processing of music. *NeuroImage*, 17, 956–966.

Koelsch, S., Gunter, T., Schröger, E., & Friederici, A. D. (2003). Processing tonal modulations: An ERP study. *Journal of Cognitive Neuroscience*, 15, 1149–1159.

Koelsch, S., Fritz, T., Schulze, K., Alsop, D., & Schlaug, G. (2005). Adults and children processing music: An fMRI study. *NeuroImage*, 25(4), 1068–1076.

Kral, A., & Eggermont, J. J. (2007). What's to lose and what's to learn: Development under auditory deprivation, cochlear implants and limits of cortical plasticity. *Brain Research Reviews*, 56(1), 259–269.

Kraus, N., McGee, T., Littman, T., & King, C. (1994). Nonprimary auditory thalamic representation of acoustic change. *Journal of Neurophysiology*, 72, 1270–1277.

Kraus, N., McGee, T., Carrell, T., King, C., Tremblay, K., & Nicol, T. (1995). Central auditory system plasticity associated with speech discrimination training. *Journal of Cognitive Neuroscience*, 7, 25–32.

Krumbholz, K., Patterson, R. D., Seither-Preisler, A., Lammertmann, C., & Lutkenhoner, B. (2003). Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. *Cerebral Cortex*, 13(7), 765–772.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. New York: Oxford University Press.

Lahav, A., Saltzman, E., & Schlaug, G. (2007). Action representation of sound: Audiomotor recognition network while listening to newly acquired actions. *Journal of Neuroscience*, 27(2), 308–314.

Lappe, C., Herholz, S. C., Trainor, L. J., & Pantev, C. (2008). Cortical plasticity induced by short-term unimodal and multimodal musical training. *Journal of Neuroscience*, 28(39), 9632–9639.

Large, E. W., & Palmer, C. (2002). Perceiving temporal regularity in music. *Cognitive Science*, 26, 1–37.

Leino, S., Brattico, E., Tervaniemi, M., & Vuust, P. (2007). Representation of harmony rules in the human brain: Further evidence from event-related potentials. *Brain Research*, 1142, 169–177.

Liégeois-Chauvel, C., Peretz, I., Babaï, M., Laguitton, V., & Chauvel, P. (1998). Contribution of different cortical areas in the temporal lobes to music processing. *Brain*, 121, 1853–1867.

Loui, P., Guenther, F. H., Mathys, C., & Schlaug, G. (2008). Action-perception mismatch in tone-deafness. *Current Biology*, 18(8), R331–R332.

Loui, P., Alsop, D., & Schlaug, G. (2009). Tone deafness: A new disconnection syndrome? *Journal of Neuroscience*, 29(33), 10215–10220.

Maess, B., Koelsch, S., Gunter, T., & Friederici, A. D. (2001). "Musical syntax" is processed in the area of Broca: An MEG-study. *Nature Neuroscience*, 4, 540–545.

Mandell, J., Schulze, K., & Schlaug, G. (2007). Congenital amusia: An auditory-motor feedback disorder? *Restor Neurology and Neuroscience*, 25(3–4), 323–334.

Margulis, E. H., Mlsna, L. M., Uppunda, A. K., Parrish, T. B., & Wong, P. C. M. (2009). Selective neurophysiologic responses to music in instrumentalists with different listening biographies. *Human Brain Mapping*, 30(1), 267–275.

Mars, R. B., Piekema, C., Coles, M. G., Hulstijn, W., & Toni, I. (2007). On the programming and reprogramming of actions. *Cerebral Cortex*, 17(12), 2972–2979.

McDermott, J. H., & Oxenham, A. J. (2008). Music perception, pitch, and the auditory system. *Current Opinion in Neurobiology*, 18(4), 452–463.

Menning, H., Roberts, L. E., & Pantev, C. (2000). Plastic changes in the auditory cortex induced by intensive frequency discrimination training. *NeuroReport*, 11, 817–822.

Milner, B. A. (1962). Laterality effects in audition. In V. Mountcastle (Ed.), *Interhemispheric relations and cerebral dominance* (pp. 177–195). Baltimore, MD: Johns Hopkins University Press.

Molholm, S., Martinez, A., Ritter, W., Javitt, D. C., & Foxe, J. J. (2005). The neural circuitry of pre-attentive auditory change-detection: An fMRI study of pitch and duration mismatch negativity generators. *Cerebral Cortex*, 15(5), 545–551.

Morillon, B., Lehongre, K., Frackowiak, R. S., Ducorps, A., Kleinschmidt, A., Poeppel, D., & Giraud, A. L. (2010). Neurophysiological origin of human brain asymmetry for speech and language. *Proceedings of the National Academy of Sciences of the USA*, 107(43), 18688–18693.

Möttönen, R., Calvert, G. A., Jääskeläinen, I. P., Matthews, P. M., Thesen, T., Tuomainen, J., & Sams, M. (2006). Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. *NeuroImage*, 30, 563–569.

Münte, T. F., Kohlmetz, C., Nager, W., & Altenmüller, E. (2001). Superior auditory spatial tuning in conductors. *Nature*, 409, 580.

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences of the USA*, 104(40), 15894–15898.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, 118(12), 2544–2590.

Ohnishi, T., Matsuda, H., Asada, T., Aruga, M., Hirakata, M., Nishikawa, M., et al. (2001). Functional anatomy of musical perception in musicians. *Cerebral Cortex*, 11(8), 754–760.

Okamoto, H., Stracke, H., Draganova, R., & Pantev, C. (2009). Hemispheric asymmetry of auditory evoked fields elicited by spectral versus temporal stimulus change. *Cerebral Cortex*, 19(10), 2290–2297.

Opitz, B., Rinne, T., Mecklinger, A., von Cramon, D. Y., & Schröger, E. (2002). Differential contribution of frontal and temporal cortices to auditory change detection: fMRI and ERP results. *NeuroImage*, 15(1), 167–174.

Overath, T., Cusack, R., Kumar, S., Von Kriegstein, K., Warren, J. D., Grube, M., et al. (2007). An information theoretic characterisation of auditory encoding. *PLoS Biology*, 5(11), 2723–2732.

Overath, T., Kumar, S., von Kriegstein, K., & Griffiths, T. D. (2008). Encoding of spectral correlation over time in auditory cortex. *Journal of Neuroscience*, 28(49), 13268–13273.

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature*, 392, 811–814.

Pantev, C., Roberts, L., Schulz, M., Engelien, A., & Ross, B. (2001). Timbre-specific enhancement of auditory cortical representations in musicians. *NeuroReport*, 12, 169–174.

Patel, A. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, 6, 674–681.

Patel, A., & Balaban, E. (2001). Human pitch perception is reflected in the timing of stimulus-related cortical activity. *Nature Neuroscience*, 4, 839–844.

Patel, A. D. (2008). *Music*, *language*, *and the brain*. New York: Oxford University Press.

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, 36, 767–776.

Penagos, H., Melcher, J. R., & Oxenham, A. J. (2004). A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *Journal of Neuroscience*, 24(30), 6810–6815.

Penhune, V. B., Zatorre, R. J., MacDonald, J. D., & Evans, A. C. (1996). Interhemispheric anatomical differences in human primary auditory cortex: Probabilistic mapping and volume measurement from magnetic resonance scans. *Cerebral Cortex*, 6, 661–672.

Penhune, V. B., Cismaru, R., Dorsaint-Pierre, R., Petitto, L. A., & Zatorre, R. J. (2003). The morphometry of auditory cortex in the congenitally deaf measured using MRI. *NeuroImage*, 20, 1215–1225.

Peretz, I., & Hyde, K. L. (2003). What is specific to music processing? Insights from congenital amusia. *Trends in Cognitive Sciences*, 7, 362–367.

Peretz, I., Brattico, E., & Tervaniemi, M. (2005). Abnormal electrical brain responses to pitch in congenital amusia. *Annals of Neurology*, 58, 478–482.

Peretz, I., Cummings, S., & Dubé, M. P. (2007). The genetics of congenital amusia (tone deafness): A family-aggregation study. *American Journal of Human Genetics*, 81, 582–588.

Peretz, I., Brattico, E., Järvenpäa, M., & Tervaniemi, M. (2009). The amusic brain: In tune, out of key, and unaware. *Brain*, 132, 1277–1286.

Perry, D. W., Zatorre, R. J., Petrides, M., Alivisatos, B., Meyer, E., & Evans, A. C. (1999). Localization of cerebral activity during simple singing. *NeuroReport*, 10, 3979–3984.

Petkov, C. I., Kang, X., Alho, K., Bertrand, O., Yund, E. W., & Woods, D. L. (2004). Attentional modulation of human auditory cortex. *Nature Neuroscience*, 7, 658–663.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time.' *Speech Communication*, 41, 245–255.

Puschmann, S., Uppenkamp, S., Kollmeier, B., & Thiel, C. M. (2010). Dichotic pitch activates pitch processing centre in Heschl's gyrus. *NeuroImage*, 49(2), 1641–1649.

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6), 718–724.

Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proceedings of the National Academy of Sciences of the USA*, 97(22), 11800–11806.

Recanzone, G. H., Guard, D. C., Phan, M. L., & Su, T. K. (2000). Correlation between the activity of single auditory cortical neurons and sound-localization behavior in the macaque monkey. *Journal of Neurophysiology*, 83(5), 2723–2739.

Rinne, T., Degerman, A., & Alho, K. (2005). Superior temporal and inferior frontal cortices are activated by infrequent sound duration decrements: An fMRI study. *NeuroImage*, 26(1), 66–72.

Rivier, F., & Clarke, S. (1997). Cytochrome oxidase, acetylcholinesterase, and NADPH-diaphorase staining in human supratemporal and insular cortex: Evidence for multiple auditory areas. *NeuroImage*, 6(4), 288–304.

Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., & Rauschecker, J. P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nature Neuroscience*, 2(12), 1131–1136.

Romanski, L. M., Tian, B., Fritz, J. B., Mishkin, M., Goldman-Rakic, P. S., & Rauschecker, J. P. (2000). Reply to "What', 'where' and 'how' in auditory cortex.' *Nature Neuroscience*, 3(10), 966.

Schlaug, G., Jancke, L., Huang, Y., Staiger, J. F., & Steinmetz, H. (1995). Increased corpus callosum size in musicians. *Neuropsychologia*, 33(8), 1047–1055.

Schlaug, G., Forgeard, M., Zhu, L., Norton, A., Norton, A., & Winner, E. (2009). Training-induced neuroplasticity in young children. *Annals of the New York Academy of Sciences* 1169, 205–208.

Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., & Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nature Neuroscience*, 5, 688–694.

Schönwiesner, M., & Zatorre, R. J. (2008). Depth electrode recordings show double dissociation between pitch processing in lateral Heschl's gyrus and sound onset processing in medial Heschl's gyrus. *Experimental Brain Research*, 187, 97–105.

Schönwiesner, M., Rubsamen, R., & von Cramon, D. Y. (2005). Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. *European Journal of Neuroscience*, 22(6), 1521–1528.

Schönwiesner, M., Novitski, N., Pakarinen, S., Carlson, S., Tervaniemi, M., & Näätänen, R. (2007). Heschl's gyrus, posterior superior temporal gyrus, and mid-ventrolateral prefrontal cortex have different roles in the detection of acoustic changes. *Journal of Neurophysiology*, 97(3), 2075–2082.

Schroeder, C., & Foxe, J. (2002). The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Cognitive Brain Research*, 14, 187–198.

Seldon, H. (1981). Structure of human auditory cortex. II: Axon distributions and morphological correlates of speech perception. *Brain Research*, 229, 295–310.

Shepard, R. N. (1982). Geometrical approximations to the structure of musical pitch. *Psychological Review*, 89(4), 305–333.

Sigalovsky, I. S., Fischl, B., & Melcher, J. R. (2006). Mapping an intrinsic MR property of gray matter in auditory cortex of living humans: A possible marker for primary cortex and hemispheric differences. *NeuroImage*, 32(4), 1524–1537.

Sluming, V., Barrick, T., Howard, M., Cezayirli, E., Mayes, A., & Roberts, N. (2002). Voxel-based morphometry reveals increased gray matter density in Broca's area in male symphony orchestra musicians. *NeuroImage*, 17, 1613–1622.

Smith, K. R., Hsieh, I.-H., Saberi, K., & Hickok, G. (2010). Auditory spatial and object processing in the human planum temporale: No evidence for selectivity. *Journal of Cognitive Neuroscience*, 22(4), 632–639.

Stewart, L., von Kriegstein, K., Warren, J. D., & Griffiths, T. D. (2006). Music and the brain: Disorders of musical listening. *Brain*, 129(10), 2533–2553.

Tervaniemi, M., Rytkönen, M., Schröger, E., Ilmoniemi. R. J., & Näätänen, R. (2001). Superior formation of cortical memory traces for melodic patterns in musicians. *Learning and Memory*, 8, 295–300.

Tervaniemi, M., Szameitat, A. J., Kruck, S., Schroger, E., Alter, K., De Baene, W., & Friederici, A. D. (2006). From air oscillations to music and speech: Functional magnetic resonance imaging evidence for fine-tuned neural networks in audition. *Journal of Neuroscience*, 26(34), 8647–8652.

Thivard, L., Belin, P., Zilbovicius, M., Poline, J., & Samson, Y. (2000). A cortical region sensitive to auditory spectral motion. *NeuroReport*, 11, 2969–2972.

Tian, B., Reser, D., Durham, A., Kustov, A., & Rauschecker, J. P. (2001). Functional specialization in rhesus monkey auditory cortex. *Science*, 292(5515), 290–293.

Tillmann, B., Koelsch, S., Escoffier, N., Bigand, E., Lalitte, P., Friederici, A., & von Cramon, D. (2006). Cognitive priming in sung and instrumental music: Activation of inferior frontal cortex. *NeuroImage*, 31, 1771–1782.

Tillmann, B., Jolicœur, P., Ishihara, M., Gosselin, N., Bertrand, O., Rossetti, Y., & Peretz, I. (2010). The amusic brain: Lost in music, but not in space. *PLoS ONE*, 5(4), e10173.

Trainor, L., McDonald, K. L., & Alain, C. (2002). Automatic and controlled processing of melodic contour and interval information measured by electrical brain activity. *Journal of Cognitive Neuroscience*, 14, 430–442.

Ungerleider, L. G., & Haxby, J. V. (1994). 'What' and 'where' in the human brain. *Current Opinion in Neurobiology*, 4(2), 157–165.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.

von Economo, C., & Horn, L. (1930). Über Windungsrelief, Maße und Rindenarchitektonik der Supratemporalfläche, ihre individuellen und ihre Seitenunterschiede. *Zeitschrift Neurologie und Psychiatrie*, 130, 678–757.

Warren, J. D., & Griffiths, T. D. (2003). Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. *Journal of Neuroscience*, 23, 5799–5804.

Warren, J. D., Uppenkamp, S., Patterson, R. D., & Griffiths, T. D. (2003). Separating pitch chroma and pitch height in the human brain. *Proceedings of the National Academy of Sciences of the USA*, 100(17), 10038–10042.

Warren, J. E., Wise, R. J., & Warren, J. D. (2005). Sounds do-able: Auditory-motor transformations and the posterior temporal plane. *Trends in Neurosciences*, 28(12), 636–643.

Warrier, C., Wong, P., Penhune, V., Zatorre, R., Parrish, T., Abrams, D., & Kraus, N. (2009). Relating structure to function: Heschl's gyrus and acoustic processing. *Journal of Neuroscience*, 29(1), 61–69.

Watanabe, D., Savion-Lemieux, T., & Penhune, V. B. (2007). The effect of early musical training on adult motor performance: Evidence for a sensitive period in motor learning. *Experimental Brain Research*, 176, 332–340.

Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, 10(4), 420–422.

Wong, P. C. M., Warrier, C. M., Penhune, V. B., Roy, A. K., Sadehh, A., Parrish, T. B., & Zatorre, R. J. (2008). Volume of left Heschl's gyrus and linguistic pitch learning. *Cerebral Cortex*, 18, 828–836.

Zacks, J. M. (2008). Neuroimaging studies of mental rotation: A meta-analysis and review. *Journal of Cognitive Neuroscience*, 20(1), 1–19.

Zarate, J. M., & Zatorre, R. J. (2008). Experience-dependent neural substrates involved in vocal pitch regulation during singing. *NeuroImage*, 40(4), 1871–1887.

Zarate, J. M., Delhommeau, K., Wood, S., & Zatorre, R. J. (2010a). Vocal accuracy and neural plasticity following micromelody-discrimination training. *PLoS ONE*, 5(6), e11181.

Zarate, J. M., Wood, S., & Zatorre, R. J. (2010b). Neural networks involved in voluntary and involuntary vocal pitch regulation in experienced singers. *Neuropsychologia*, 48(2), 607–618.

Zatorre, R. J. (1985). Discrimination and recognition of tonal melodies after unilateral cerebral excisions. *Neuropsychologia*, 23, 31–41.

Zatorre, R. J. (1988). Pitch perception of complex tones and human temporal-lobe function. *Journal of the Acoustical Society of America*, 84(2), 566–572.

Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, 11, 946–953.

Zatorre, R. J., & Gandour, J. T. (2007). Neural specializations for speech and pitch: Moving beyond the dichotomies. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 363, 1087–1104.

Zatorre, R. J., Evans, A. C., & Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *Journal of Neuroscience*, 14(4), 1908–1919.

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002a). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Science*, 6, 37–46.

Zatorre, R. J., Bouffard, M., Ahad, P., & Belin, P. (2002b). Where is 'where' in the human auditory cortex? *Nature Neuroscience*, 5, 905–909.

Zatorre, R. J., Bouffard, M., & Belin, P. (2004). Sensitivity to auditory object features in human temporal neocortex. *Journal of Neuroscience*, 24(14), 3637–3642.

Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: Auditory-motor interactions in music perception and production. *Nature Reviews Neuroscience*, 8(7), 547–558.

Zatorre, R. J., Halpern, A. R., & Bouffard, M. (2010). Mental reversal of imagined melodies: A role for the posterior parietal cortex. *Journal of Cognitive Neuroscience*, 22, 775–789.

Chapter 11
Multisensory Role of Human Auditory Cortex

Virginie van Wassenhove and Charles E. Schroeder

11.1 Introduction

Multisensory research is at an early stage of inquiry and provides fascinating evidence questioning the long held view that sensory modalities are independent analytical pathways. Auditory cortex is a prime example of cortical area that can often be modulated by inputs coming from different sensory and motor modalities.

Multisensory research has largely benefited from the systematic description of neural populations in the multisensory layers of the superior colliculus (Stein & Meredith, 1993). Three major rules of multisensory integration have been utilized to specify multisensory interactions at many levels of cortex. In the *spatial rule*, the degree to which multisensory inputs overlap with the unisensory spatial receptive fields of multisensory neurons determines whether supra- or sub-additive responses will be observed: if the stimuli are in spatial register, neural responses are supra-additive; otherwise, sub-additive responses can be observed. Likewise in the *temporal rule*, the degree to which multisensory inputs temporally overlap with the unisensory receptive fields of multisensory neurons determines the observed response type, namely supra- or sub-additive when stimuli are or not in temporal register, respectively (Benevento et al., 1977; Meredith et al., 1987). In the *inverse effectiveness rule*, the responses of multisensory neurons to multisensory stimulations are most effective (enhanced) when unisensory stimulations are least effective.

V. van Wassenhove (✉)
CEA DSV.I^{2}BM.NeuroSpin, Cognitive Neuroimaging Unit (INSERM U992),
Bât 145 - Point Courrier 156, Gif s/Yvette F-91191, France
e-mail: Virginie.van-Wassenhove@cea.fr; Virginie.van.Wassenhove@gmail.com

C.E. Schroeder
Professor of Psychiatry, Cognitive Neuroscience & Schizophrenia Program,
Nathan S. Kline Institute for Psychiatric Research, 140 Old Orangeburg Road,
Orangeburg, NY 10962, USA
e-mail: Schrod@nki.rfmh.org

D. Poeppel et al. (eds.), *The Human Auditory Cortex*, Springer Handbook
of Auditory Research 43, DOI 10.1007/978-1-4614-2314-0_11,

These three canonical rules have subsequently been applied to human neuroimaging data including functional magnetic resonance imaging (fMRI), electroencephalography (EEG), and magnetoencephalography (MEG). These rules have also inspired a range of paradigmatic approaches in which the introduction of noise (as spatial misalignment and/or temporal desynchronization) assesses the constraints of multisensory integration.

Crucially, a growing body of evidence has suggested that multisensory neurons are not the only site for multisensory cross-talks: sensory cortices are implicated in a direct or indirect manner in multisensory integration. Several mechanistic accounts for these intersensory interactions have been put forward. This chapter reviews and discusses the functional implication of auditory cortical regions in multisensory integration: Section 11.2 focuses on the neurophysiology of auditory cortex in monkeys and associated findings relevant to multisensory processing; Section 11.3 reviews psychophysical and neuroimaging findings in which auditory cortex has been implicated; and Section 11.4 specifically focuses on auditory–visual (AV) speech.

11.2 Nonhuman Neuroanatomy and Neurophysiology

Following up on early indications of multisensory interactions in low-level auditory cortices in humans (Calvert et al., 1997; Levänen et al., 1998; Giard and Peronnet, 1999) and monkeys (Schroeder et al., 2001), several groups have conducted intensive investigations of nonauditory influences on the auditory cortices (Kayser et al., 2005; Bizley et al., 2007; Lakatos et al., 2007) and of the underlying anatomical circuits (Budinger et al., 2006; Hackett et al., 2007a, 2007b; Cappe et al., 2009) in nonhuman species. The recent surge of investigation merges with a complex history of findings on somatosensory representations in the region of posterior auditory cortex. Starting with Leinonen et al. (1980) and Robinson and Burton (1980), recent studies confirm somatosensory input into the caudomedial (CM) auditory area and suggest multiple somato-auditory fields in the superior temporal plane. Likewise, examination of the partial maps in Krubitzer et al. (1995) suggests the possibility of two or more body maps in the superior temporal plane. At this point, the number of body maps, as well as the content of the map(s) in the posterior superior temporal plane, is unclear. The following section considers anatomical source(s) of multisensory convergence in auditory association cortex and other locations in which convergence occurs.

11.2.1 Anatomical Circuits for Multisensory Convergence in Low-Level Cortices

The known connectivity patterns of the primate brain provide a number of routes by which sensory inputs can converge in a given cortical area. Somatosensory and visual inputs to putatively unisensory auditory cortices, as shown by recent tract tracing studies, provide well-studied examples illustrating this point (Fig. 11.1).

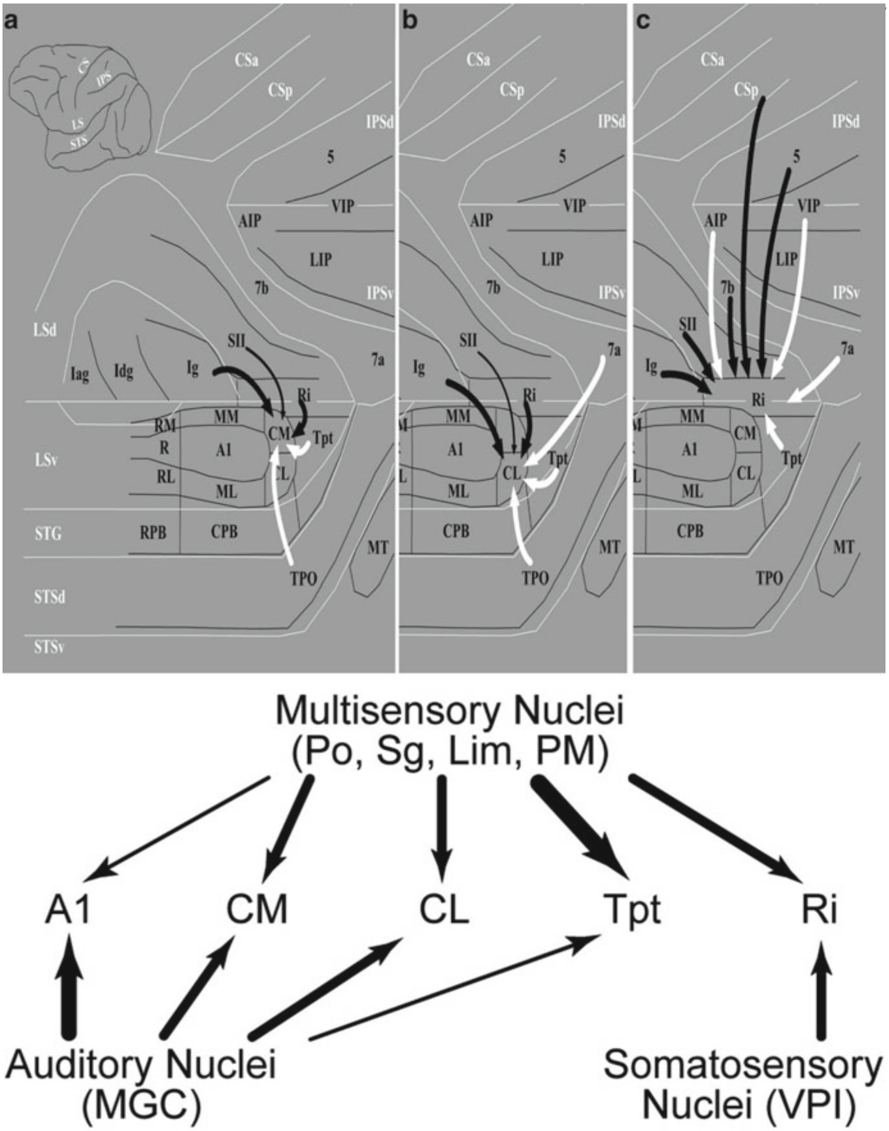


Fig. 11.1 (Top) Summary of potential sources of somatosensory input to the caudal auditory belt areas. Black arrows show input that are presumed to carry mainly somatosensory information. White arrows show inputs from areas that are known to be sites of multisensory integration. Thin arrows show connections that were weak and inconsistent. (**a**) Projections to area CM (caudomedial belt cortex) include inputs from Ri (retroinsular area), Ig (granular insula), and perhaps weakly from SII (second somatosensory area) that are presumed to be mainly somatosensory inputs. Inputs were also seen from areas Tpt and TPO that are both thought to be sites of multisensory integration. (**b**) Injections of CL (caudolateral belt cortex), and also lateral Tpt (temporoparietal area) showed inputs very similar to those of CM, but in addition there were projections from area 7a in caudal inferior parietal cortex. (**c**) Projections to area Ri were from several somatosensory and multisensory areas in parietal cortex. (Adapted from Smiley et al., 2007.) (Bottom) Summary of thalamocortical inputs to A1, CM, CL, Tpt, and Ri. Heavier arrows indicate denser projections. Auditory areas receive the densest inputs from the MGC and variable projections from the multisensory nuclei. Tpt receives the densest multisensory inputs and modest auditory projections. Ri has uncertain auditory inputs, and stronger inputs from multisensory and somatosensory nuclei. (Adapted from Hackett et al., 2007a.)

Depicted is a summary of the potential sources of somatosensory input to the posterior regions of auditory cortex adjacent to A1 demonstrated in a recent series of fluorescent tracer injection studies in macaque monkeys. From these it is clear that auditory cortices receive extensive projections from classic nonauditory areas, any of which could serve as a substrate of auditory influences over auditory processing. "Suborning" of activity in unisensory cortex by feedback-mediated input from another sensory modality is an intriguing phenomenon. The auditory cortices appear to contain the substrates for very early multisensory integration. In view of the multiple possible circuits for multisensory convergence in auditory cortex, or, for that matter any sensory area, it is in some respects remarkable that there is enough sensory segregation to promote unisensory representation at early cortical processing stages. Obviously, under normal circumstances, there is a strong bias in favor of specific patterns of unimodal representation and processing in lower order cortical areas. However, the striking findings in the congenitally deaf human (Levänen et al., 1998), along with the earlier work on developmental sensory rewiring in ferrets (Pallas et al., 1990; Roe et al., 1990; Sur et al., 1990), indicate that there is a great potential for altering this bias during development.

11.2.2 Physiological Manifestations of Multisensory Convergence in Low-Level Cortices

Current models of visual (Felleman & Van Essen, 1991), auditory (Rauschecker et al., 1997), and somatosensory (Burton & Sinclair, 1996) organization clearly illustrate that, using feedforward, feedback, or lateral projection circuits, inputs from any sensory source can be routed to nearly any cortical region, and even to most subcortical regions. At this time, there is a growing understanding of the brain circuits that promote multisensory convergence, both from a *structural and a functional* perspective. Structurally, feedforward and feedback inputs into a cortical area can be distinguished by the laminar pattern of axon terminations (Rockland & Pandya, 1979; Felleman & Van Essen, 1991). Feedforward input terminations are concentrated in and near lamina 4, while feedback projections largely exclude lamina 4, terminating either in the supragranular laminae, or in a "bilaminar" pattern above and below lamina 4. Lateral projections have a "columnar" pattern, terminating without any particular laminar focus (Felleman & Van Essen, 1991). These laminar termination patterns make predictions about laminar patterns of sensory responses, which can be addressed physiologically, as discussed in the next section.

11.2.2.1 Functional Manifestations of Feedforward Convergence

Recording of laminar response profiles with linear array multielectrodes is an established means of defining the functional correlates of feedforward inputs in visual (Givre et al., 1994; Schroeder et al., 1998; Mehta et al., 2000b), somatosensory

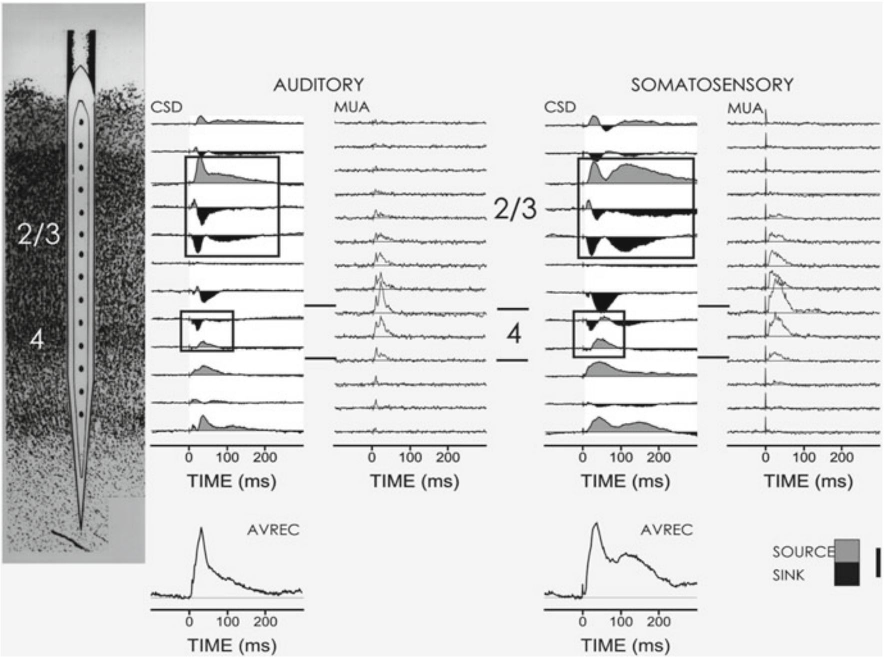


Fig. 11.2 Laminar current source density (CSD) and concomitant multiunit activity (MUA) profiles (each an average of 100 trials) sampled using a multielectrode with a linear array of recording contacts (150-mm spacing) straddling auditory area CM. Auditory (left) and somatosensory (right) responses were elicited in this site by binaural 65-dB clicks and contralateral median nerve stimuli, respectively. Downward CSD deflections (dark) signify net extracellular current sinks (inward transmembrane current); upward deflections (stippled) indicate net extracellular current sources (outward current). MUA histograms are obtained by full-wave rectification and averaging of the high-frequency activity at each electrode contact. The boxes circumscribe CSD configurations that reflect the initial excitatory response at the depth of lamina 4, and the subsequent excitation of the pyramidal cell ensembles in laminae 2/3. Averaged rectified current flow (AVREC) waveforms (bottom) provide a condensed representation of the temporal activation pattern collapsed across laminae and current flow (source/sink) direction. Scale bar (bottom right) 1.4 mV/mm^2 for CSD, 0.1 mV/mm^2 for AVREC, and 1.6 mV for MUA. (Adapted from Schroeder et al., 2001.)

(Peterson et al., 1995; Schroeder et al., 1995, 1997), and auditory (Steinschneider et al., 1994, 1998; Schroeder & Foxe 2002) cortices. Such recordings in areas of multisensory convergence with linear array multielectrodes (Schroeder et al., 2001) provide an opportunity to investigate the functional correlates of these input patterns. The laminar profile of auditory response in auditory association cortex (e.g., Fig. 11.2) has the pattern predicted by the anatomy of feedforward input: the initial response is centered on lamina 4, and followed by responses in the extragranular laminae. Feedforward auditory inputs to the region in question, CM auditory cortex, are well established (Kosaki et al., 1997; Hackett et al., 1998).

In the same location, the overall timing and the laminar activation sequence for a convergent somatosensory input are nearly identical to the timing and sequence of

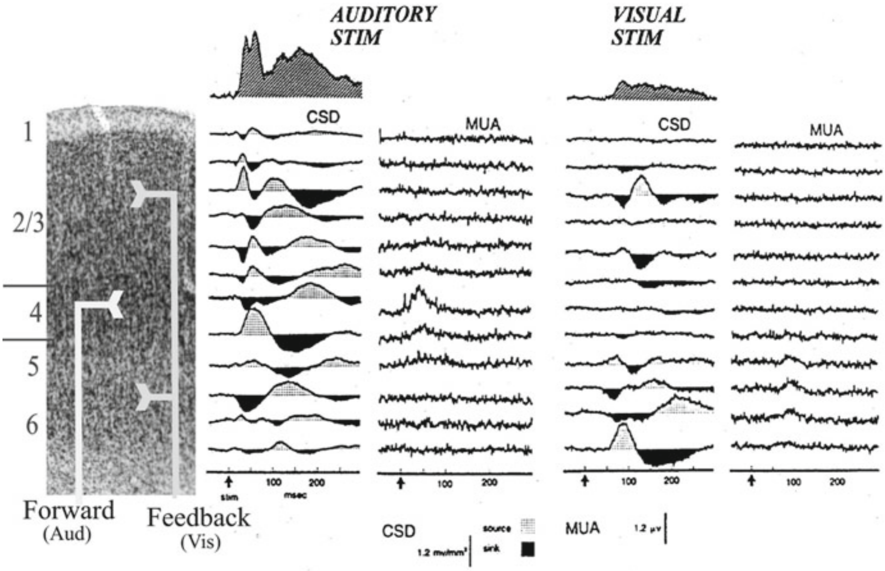


Fig. 11.3 Laminar CSD and MUA profiles evoked by auditory (left) and visual stimuli (right) and recorded from one site in auditory association cortex, poster lateral to A1 cortex. Intercontact spacing on the multielectrode was 150 mm. Each tracing represents an average of 100 stimulus-evoked responses. Those on the left represent the averaged responses to binaural 65-dB clicks. Those on the right were elicited by intense binocular light flashes (10 ms 5 duration, 7.8310 lux intensity). The CSD configurations reflect the initial excitatory response at the depth of lamina 4 (auditory profile), as opposed to above and below lamina 4 (visual profile). At the extreme left is a diagram depicting the laminar pattern of termination for feedforward inputs from auditory and feedback visual systems. (Adapted from Schroeder and Foxe, 2002.)

the auditory input. This suggests that the somatosensory input, like the auditory input, is conveyed by a feedforward projection. The source of the somatosensory input is unclear at this time, because the anatomical interface between the auditory and somatosensory areas of the lateral sulcus region is not clearly delineated.

11.2.2.2 Functional Manifestations of Associative Convergence

The laminar activity analysis findings (e.g., Schroeder & Foxe, 2002; Lakatos et al., 2007) also point to an "associative" convergence in auditory cortex, that is one that is mediated by feedback or lateral corticocortical inputs or by extralemniscal thalamic inputs; we first illustrate a nonauditory "associative" type of input profile that, based on laminar pattern and onset timing, is likely to be mediated by either feedback or lateral type cortical inputs (see Fig. 11.3). As was shown earlier (Fig. 11.2), auditory inputs have the characteristics of a feedforward anatomical projection: that is, with initial activation centered on lamina 4. This is typical throughout the core and belt regions of auditory cortex.

The visual input profile, in contrast, has a multilaminar pattern with responses beginning simultaneously in the supra- and infragranular laminae. This is the physiological pattern predicted by the anatomy of feedback and lateral projections (Rockland & Pandya, 1979; Felleman & Van Essen, 1991). Another point of contrast between the colocated auditory and visual response profiles concerns response timing. Visual response latency (~50 ms) is considerably longer than the auditory response latency (~11 ms). The large timing difference between convergent visual and auditory inputs to a single location contrasts with the lack of any corresponding timing difference between convergent somatosensory and auditory inputs to single auditory cortical locations (Figs. 2 and 5). With proper analysis and interpretation, these latency data can help to identify input sources and possibly also to predict the characteristics of multisensory interactions.

11.2.2.3 Functional Manifestations of Extralemniscal Thalamic Inputs

Projections to auditory cortex from a variety of extralemniscal thalamic nuclei, including so-called nonspecific and multisensory nuclei as well as from nuclei devoted to other modalities such as the ventral–posterior inferior nucleus (VPI), and even higher order thalamic regions (such as pulvinar) provide additional potential mechanisms for multisensory convergence. These parallel the main or “core” thalamic projections into the neocortex, and although they generally target cortex appropriate for the sense modality from which they originate, there are exceptions. For example, the ventral–posterior complex, which mainly carries somatosensory inputs, contains koniocellular neurons that project sparsely to posterior auditory cortex.

11.2.2.4 Signature of Driving versus Modulatory Input

Many of the effects of nonauditory input into auditory cortex are likely to be in the form of modulatory as opposed to driving inputs (see earlier). In attempting to measure these often subtle effects, analysis of concomitant local field potential and multiunit activity across the cortical laminae allows functional differentiation between *driving and modulatory inputs* as well as the direction of input (feedforward vs. feedback) (Lipton et al., 2006; Chen et al., 2007; Lakatos et al., 2007). Comparison of the qualitative and quantitative patterns of the auditory and somatosensory inputs to a given site in A1 (Fig. 11.4) illustrates this point. Unlike the auditory event–related response, the somatosensory-related current source density (CSD) response is much less intense, and there is no consistent phasic multiunit activity (MUA) correlate. Therefore, the somatosensory input by itself does not appear “effective,” in that it does not drive detectable action potentials threshold in local neurons. Thus, rather than conveying specific information, the somatosensory input appears to be “modulatory.” Compounding this observation is the difference in timing and laminar profile noted earlier.

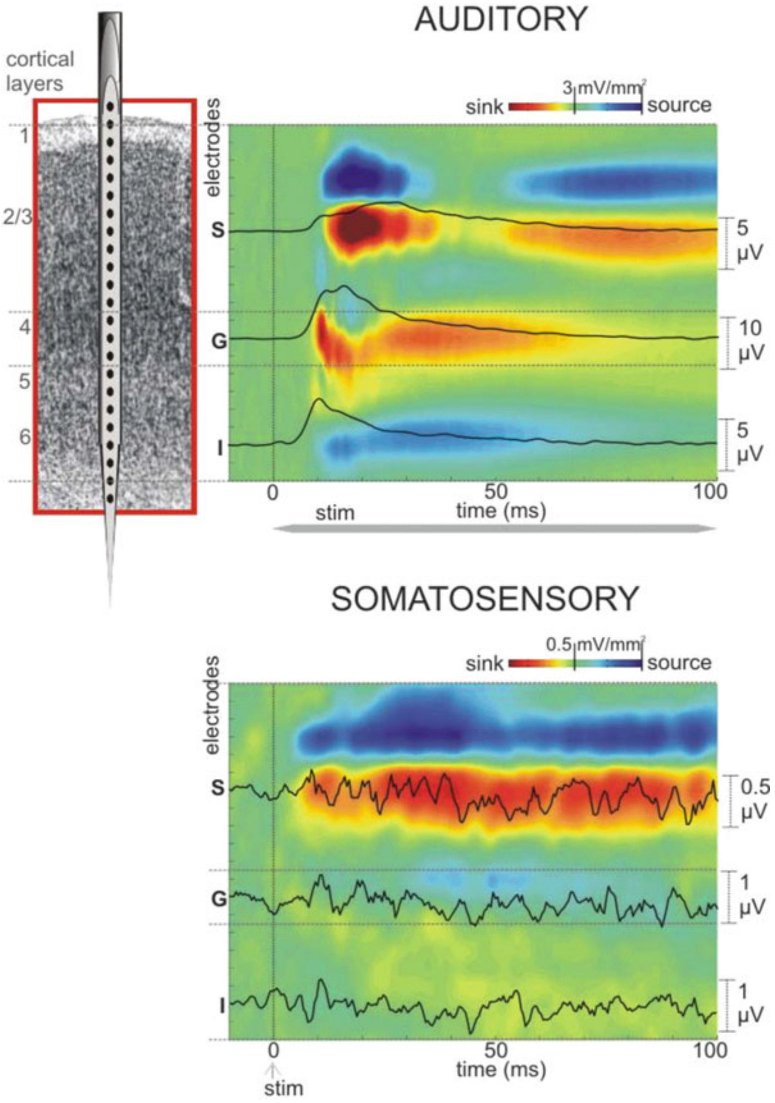


Fig. 11.4 (A) Field potentials (used to calculate the CSD) and MUA were recorded concomitantly with a linear-array multicontact electrode positioned to sample from all cortical layers. Laminar boundaries were determined based on functional criteria. Color maps show the laminar profiles of a representative CF tone and a somatosensory stimulus-related averaged CSD (98 and 95 sweeps, respectively), recorded in the same location. Current sinks (net inward transmembrane current) are red and current sources (net outward transmembrane current) are blue. Based on their largest amplitude in the auditory CSD, one electrode was selected in each layer (S, G, and I) for quantitative analysis. Overlaid traces show MUA in the selected channels. (Adapted from Lakatos et al., 2007.)

11.2.3 Constraints Imposed by Input Timing

Studies in macaque auditory cortex can directly and accurately resolve input timing in a way that can significantly enhance our understanding of the dynamics of auditory cortical processing in humans.

11.2.3.1 Auditory Response Timing in Macaque Auditory Cortex

In accord with earlier studies (Recanzone et al., 2000; Brosch et al., 2005; Kajikawa et al., 2005), Lakatos and colleagues (2005) showed median-onset response latencies to characteristic frequency (CF) tones at 9 ms in A1 and 12 ms in auditory belt regions. Response timing to broad-band noise (BBN) was similar to CF response in A1 (median = 8.5 ms) but significantly shorter in the belt region (median = 7 ms). Our estimated A1 latencies to broadband noise accord with values reported by Steinschneider and colleagues (Steinschneider et al., 1992), which showed that the click-evoked initial auditory evoked potential (AEP) component in A1 began approximately 5.5 ms and peaked at approximately 8.5. Combined, these studies support a correspondence between simian response timing and waveform features in the simian system and reports in human A1 (Celesia, 1968; Liégeois-Chauvel et al., 1991).

11.2.3.2 Visual and Somatosensory Response Timing in Auditory Cortex

Using the same techniques, we have tracked the latency of visual and somatosensory responses across numerous levels of sensory processing have been tracked in awake-behaving monkeys to the presentation of high-intensity stimuli (Schroeder & Foxe, 2002; Schroeder et al., 2001; Chen et al., 2007). In the *visual responses*, initial excitation in V1 occurred at approximately 25–30 ms poststimulation, with variable but significant lags across four successive cortical stages. Whereas the timing lag across stages is similar to that in the auditory system, absolute latency is obviously much longer, and an important factor to consider in AV interactions. Of particular note are the latencies in the visual areas in the dorsal bank of the superior temporal sulcus (STSd; corresponding mainly to the superior temporal polysensory areas, which are widely considered to be likely sources of visual feedback input to auditory cortex); these are in the range of 30–35 ms across successive areas. The findings are in close agreement with early studies of onset latency in layer 4 of V1, which reported minimum response latencies of 20–31 ms (Maunsell & Gibson, 1992). A comprehensive single-unit study of the magno- and parvocellular systems showed longer latencies across areas (mean V1 = 66 ms; mean V2 = 82 ms; mean V4 = 104 ms), but also reported that latencies as short as 34 ms could be recorded in V1 when the units were isolated to layer 4 (Schmolesky et al., 1998). Although some variability in onset latency is expected across studies, owing to differences in stimulus parameters, recording techniques, and onset criteria, the minimum response

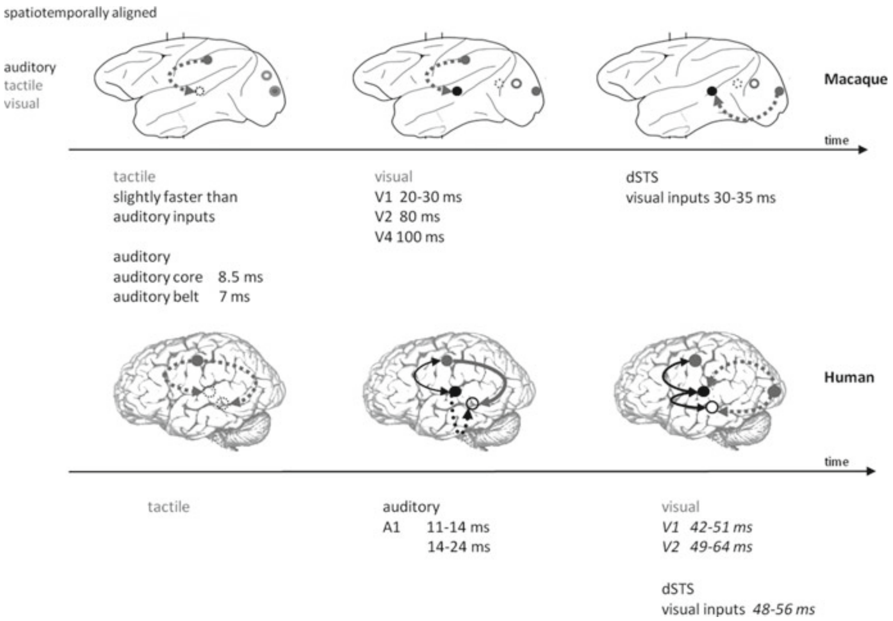


Fig. 11.5 Summary comparison of neural latencies observed in sensory and multisensory cortices for a hypothetical source of tactile, visual, and auditory events. Neural latencies reported for humans are partly extrapolated from values recorded in monkeys and partly rest on electrophysiology literature (see text). Somatosensory and auditory cortical latencies are much shorter (around 15 ms) than those in visual cortices (nearly 50 ms), including auditory responses to tactile inputs. Convergence of multisensory information onto the superior temporal sulcus can be recorded within 50 ms

latencies observed in visual cortex seems to have a high degree of reliability in the literature. Under optimal auditory latency conditions (i.e., loud stimuli, originating close to the ear), *somatosensory responses* (whether they are modulatory or driving responses) are slightly faster than auditory responses at each level of the system. As mentioned previously, in A1, somatosensory responses are slightly faster than auditory responses, though they appear to reflect a modulatory rather than a driving type of input. It is likely that somatosensory input to auditory association (belt) areas is accomplished via afferent feedforward connections with somatosensory cortex whereas somatosensory input to primary auditory cortices may be mediated by extralemniscal afferents, for example, via the nonspecific/ multisensory thalamic systems (Hackett et al., 2007a, 2007b).

11.2.3.3 Extrapolation from Monkey to Human Latencies

Using the “3/5 rule”, namely, monkey latencies tend to be about 3/5 of the corresponding values in humans (Fig. 11.5), we can extrapolate these values to corresponding “predicted” latencies in humans (Schroeder et al., 2004). With this rule, we would predict: (1) primary auditory cortical latencies of 11–14 ms and (2) primary visual cortical latencies of 42–51 ms in humans. We also predict auditory cortical belt regions in humans to have click/noise latencies identical to those in A1, and higher-order auditory cortical pure tone latencies of 14–24 ms in humans. For the second-order visual (extrastriate) cortices that are regarded as sources of direct crossing inputs to auditory cortices, predicted visual latencies would be 49–64 ms, and for the STS areas that may provide feedback input to auditory cortices, predicted visual latencies would be 48–56 ms. These predictions for primary visual and auditory cortical latencies are generally supported in the literature (e.g., Steinschneider et al., 1992; Musacchia & Schroeder, 2009).

11.2.3.4 Implications of Response Timing

There is already evidence (see earlier) that somatosensory inputs arrive in auditory cortex as early or earlier than auditory inputs (Schroeder et al., 2001; Lakatos et al., 2007). The lag between auditory and visual response indicates that precisely synchronized (simultaneous) auditory and nonauditory stimulation will produce arrival times in auditory cortex with at least as great an audiovisual lag. However, the multisensory integration process can apparently overcome this problem. *First*, although estimates of the "temporal window of integration" vary widely as a function of sensory and cognitive factors, under some circumstances the window can be as wide as 250 ms (Massaro et al., 1996; Munhall et al., 1996; Miller & D'Esposito, 2005; Conrey & Pisoni, 2006; van Wassenhove et al., 2007). Moreover, based on findings by Lakatos and colleagues (2007), multiple windows of integration can coexist, which correspond to the periods of prominent oscillatory cycles in auditory cortex, particularly those in delta, theta, and gamma bands (see Section 11.4.2.). The factors that control the effective window(s) of integration are a current focus of research efforts (Schroeder et al., 2008). *Second*, visual stimulation often precedes sound onset in natural conditions (e.g., the sight of a hammer swinging precedes the sound of the strike, and the lips and mouth form the shape of the word before phonation occurs). Normative lags between lip movements and sound onset recently catalogued by Ghazanfar and colleagues (Chandreskaran et al., 2009) indicate that in agreement with earlier more anecdotal reports (van Wassenhove et al., 2005; Schroeder et al., 2008) there is a typical visual to auditory lag of 150–200 ms in face-to-face communication. Thus, in many common circumstances, including conspecific communication, visually driven input will typically arrive in auditory cortex at or before the arrival time of auditory input, facilitating integration with afferent auditory processing. Visual lead, relative to auditory input, is likely a *requirement* for very early AV interaction. These hypotheses have been supported recently by EEG and MEG data, as discussed in Section 11.4.3.

11.3 Contributions of Auditory Cortex to Multisensory Perception

Current models of multisensory integration make the distinction between pre- and postattentional multisensory integration (also referred to as "early" and "late" integration, respectively). In addition to the timing of interactions, dichotomies are also drawn from the anatomical locus of integration and the implication of subcortical structures, sensory cortices (auditory, visual, somatosensory) and/or multisensory areas (including temporal, parietal, and frontal cortices). Numerous neuroimaging and neurophysiological studies have shown that multisensory effects coexist at different levels of the sensory hierarchy. This implies that each region is likely to have a precise function in a given multisensory process and within the perceptual and cognitive systems under scrutiny.

11.3.1 Auditory Cortex Contribution to Tactile and Auditory–Tactile Perception

The influence of sound on tactile perception was originally noted by von Schiller (1932), who suggested that the presentation of repeated auditory noise bursts could affect the perception of tactile roughness. In the "parchment skin illusion," the presentation of a 1- to 2-Hz auditory rate with variable intensity or high-frequency content reliably changes the tactile perception of rubbing hands (Jousmäki & Hari, 1998). This illusion is time dependent as its strength decreases with increased asynchrony between tactile and auditory stimuli (Jousmäki & Hari, 1998; Guest et al., 2002).

Recent evidence supports a frequency-dependent coupling of auditory and tactile perception (Yau et al., 2009). Participants decided which of two tactile stimuli was higher in frequency. During a trial, an auditory distracter (a pure tone or a bandpass noise) was presented at the same time as the comparison stimulus. Auditory distracters affected the perceptual judgment of tactile stimuli in a frequency-dependent manner: the perception of tactile stimuli was primarily biased by low-frequency (<500 Hz) sounds. The authors predicted that the auditory area CM to primary auditory cortex (A1) may be the preferred site for such auditory–tactile interaction in agreement with neurophysiological findings (Schroeder et al., 2001; Schroeder & Foxe, 2002; Fu et al., 2003).

The activation of area CM during auditory–tactile interactions has been described with fMRI and MEG. For instance, the response of auditory cortex to a 1-kHz tone is modulated by median nerve stimulation as early as 50 ms poststimulus onset (Foxe et al., 2000) and tactile stimulation modulates the response in contralateral auditory cortex regardless of spatial alignment with auditory inputs (Murray et al., 2005). An fMRI study has suggested that area CM was the most likely site of tactile modulation in humans (Foxe et al., 2002).

Using MEG, Caeteno and Jousmäki (2006) reported the activation of contralateral primary somatosensory cortex (SI), bilateral secondary somatosensory cortices (SII), and auditory cortex to the presentation of 200-Hz vibrotactile stimuli. Specifically, the SI response occurred at approximately 60 ms poststimulus onset followed by a bilateral SII and a transient auditory response at 100–200 ms. A sustained field from 200 to 700 ms was observed in auditory cortex after stimulation, in agreement with subsequent fMRI findings showing activation of the posterior auditory parabelt areas during the presentation of tactile stimuli (Schürmann et al., 2006). Conversely, initial MEG studies looking at auditory–tactile interactions suggested an auditory modulation of SII activation (Lütkenhöner et al., 2002). As observed with monkey neurophysiology, tactile stimulations likely have modulatory rather than driving effects on auditory cortex response.

11.3.2 Auditory Cortex Contribution to Visual and Auditory–Visual Perception

11.3.2.1 Ventriloquism and Spatial Capture Effects

Early findings on the visual biasing of auditory perception ("visual capture" effects; Stratton, 1897; Thomas, 1941) are now referred to as "ventriloquism effects" (Howard & Templeton, 1966; Pick et al., 1969; Bertelson & Radeau, 1981); these effects pertain to the perceptual displacement of a sound location toward that of a concomitant visual event (Fig. 11.6A). The localization of auditory events can also be captured by tactile inputs, an auditory–tactile analogue to the AV ventriloquism (Caclin et al., 2002). These effects are very robust, but it is unclear at which auditory processing stage AV integration occurs: psychophysical findings compete between a preattentional biasing of auditory processing (Bertelson et al., 2000; Vroomen et al., 2001) and a "top-down" influence of spatial attention (e.g., Driver & Spence, 1998a,b, 2004).

To characterize the temporal and spatial specificities of the ventriloquism effects, Bonath et al. (2007) used EEG and fMRI. In their study, participants reported where (left, right, middle) a sound was originating from while light-emitting diodes (LEDs) at different spatial locations lit up congruently (no illusion) or incongruently (illusion inducing) with the veridical location of the sound. The EEG analysis revealed a N260 suggesting a late (postattentional) modulation of auditory cortex in this illusion.

Several results complicate the role of auditory cortex in ventriloquism. First, auditory cortex activation is not always observed in relation to the ventriloquism effect. For instance, Macaluso et al. (2004) used an orthogonal design to test the effects of spatial and temporal misalignments with AV speech stimuli. Although different brain regions showed sensitivity to temporal (ventral stream, STS) and spatial (dorsal stream, parietal lobule) misalignments, none of the conditions specifically affected the response in auditory cortices. Similarly, Bischoff et al. (2007) used fMRI to test the misalignments in space and time of nonmeaningful AV stimuli: in the synchronous but spatially misaligned AV trials eliciting a ventriloquist illusion, no specific activation of auditory cortices was observed in relation to the illusory effects.

In contrast, evidence of *preattentional* modulation of auditory cortices has been put forward using the mismatch negativity (MMN) paradigm (Näätänen, 1995) (see Alain and Winkler, Chapter 4, for explanation). Using the MMN paradigm, Stekelenburg et al. (2004) showed that a mismatch response could be elicited when deviant stimuli consisting of spatially misaligned AV events were contrasted with standard stimuli consisting of spatially aligned AV events. Importantly, the standard and deviant sounds were identical and the sole difference resided in the spatial location of the paired visual events. The elicited mismatch response was observed within

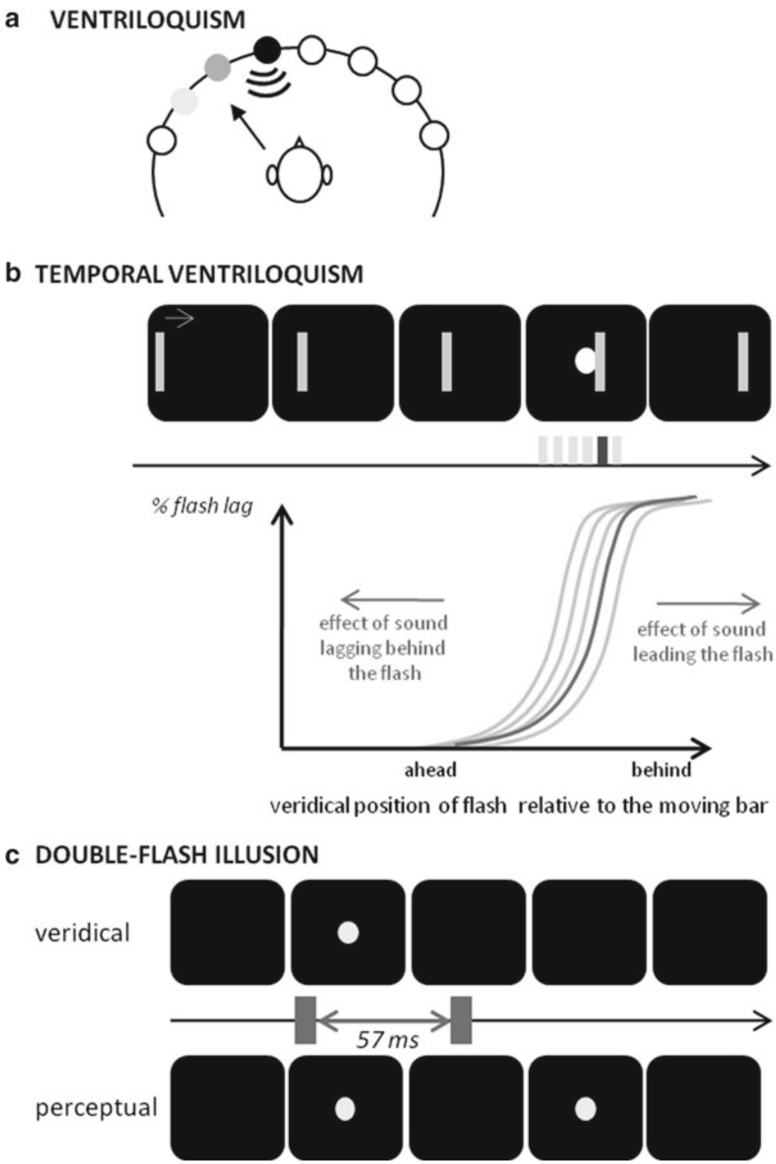

a VENTRILOQUISM
b TEMPORAL VENTRILOQUISM
% flash lag
effect of sound lagging behind the flash
effect of sound leading the flash
ahead
behind
veridical position of flash relative to the moving bar
c DOUBLE-FLASH ILLUSION
veridical
57 ms
perceptual

Fig. 11.6 Examples of multisensory integration using transient auditory and visual stimuli. (**a**) The classic experimental setup for testing *spatial ventriloquism* effects consists of an array of loudspeakers and LEDs located at a specific distance from the participant. When an LED is flashed at the same time as a sound but at a different spatial location, the sound will be perceived closer to the visual source than it actually is. (**b**) The *temporal ventriloquism* consists in biasing the temporal locus of a visual source by a sound. In this particular setup (Vroomen & de Gelder, 2004), a bar is moving across the screen toward the right at a constant speed when a disc is flashed on the display. The participant is asked to report whether the disc led or lagged the moving bar. When a transient sound is played before or after the disc, the subjective temporal locus of the disc is captured in the temporal direction of the sound. (**c**) In the *double-flash illusion* (Shams et al., 2000), a disc is briefly displayed on a screen and accompanied with two or more transient sounds. Participants are asked to report how many flashes they perceive; the number of transient sounds surrounding the brief flash strongly biases the report of the participants. When two sounds are played with a single flash, participants often report seeing two flashes (see text)

200 ms of stimulus onset, suggesting that AV interaction occurred preattentively and that sources contributing to the elicitation of the auditory MMN were sensitive to the AV spatial misalignment. As an additional control, a deviant sound placed at a different location was contrasted with a standard sound in the same location as tested with the AV standard events. In this control, the auditory MMN was found to be identical to the MMN elicited with the AV deviants, that is, *an MMN resulting from the veridical displacement of a sound is comparable to the MMN elicited by a visually induced displacement of the same sound.*

Ventriloquism aftereffects in auditory localization can also be observed as a result of adaptation to audiovisual disparities (Canon, 1970; Radeau, 1974; Lewald et al., 2001). For instance, Recanzone (1998) showed that after 30 minutes of exposure to a visual event located 8° apart from the source of a concomitant sound, participants experienced a shift in their acoustic space of approximately the same amount of disparity (~7°). This ventriloquist after-effect is observed when exposure uses different sound frequencies (here, 0.75 and 3 kHz); importantly however, no transfer can be obtained, that is, exposure to 0.75 kHz (3 kHz) does not lead to an after-effect at 3 kHz (0.75 kHz). The frequency-dependency of the ventriloquist after-effect led to the hypothesis of a potential involvement of A1 neurons, and similar after-effects were observed in monkeys (Woods & Recanzone, 2004). Despite this appealing and controversial hypothesis, no empirical testing has yet been provided.

Very recently, a new hypothesis has been put forward after the characterization of ventriloquism after-effects in two neurological populations: hemianopic and neglect patients (Passamonti et al., 2009). In this study, the exposure to spatially misaligned AV events led to after-effects in neglect but not in hemianopic patients when the exposure targeted the affected field. In neglect patients, exposure to spatially congruent AV events improved subsequent auditory localization performance (more so at the location of the adaptation). The authors proposed the implication of two possible routes in these effects: a first geniculostriatal corrective route and a second collicular–extrastriatal route involved in correction error, which may entail cross-modal plasticity as early as the inferior colliculus.

In spite of being a classic phenomenon, the ventriloquism effects are far from being fully understood. To date, no data allow clarifying in a systematic fashion the interplay between early visual influences and late attentional modulations of auditory cortex responses in spatial perception.

11.3.2.2 Temporal Coincidence

Visual capture effects such as the ventriloquism effect are not restricted to spatial phenomena but also extend to the temporal dimension (Radeau & Bertelson, 1987; Fendrich & Corballis, 2001; Soto-Faraco et al., 2002). For instance, the flash-lag effect (FLE) consists in perceiving a static visual event as lagging behind a moving visual bar when it is being flashed while the bar is in motion (Nijhawan, 1994). This effect has been reported in AV contexts (Alais & Burr, 2003; Hine et al., 2003; Vroomen & de Gelder, 2004). Using a similar paradigm (Vroomen & de Gelder, 2004; Stekelenburg & Vroomen, 2005), the timing of a transient auditory input with respect to the visual flash influences participants' report on *when* the flash occurred with respect to a moving bar; the FLE is smaller if the sound precedes the flash, and larger if the sound follows it (Fig. 11.6B). Using EEG, Stekelenburg and Vroomen (2005) tested the temporal capture of vision by audition by using this paradigm. The AV effect was associated with increased or decreased amplitude of the early visual evoked responses when the sound preceded or followed the flash, respectively.

Temporal capture phenomena can entail pure temporal aspects, such as the rate of presentation. A classic finding is the flicker–flutter illusion (Ogilvie, 1956; Gebhard & Mowbray, 1959; Shipley, 1964), in which the rate of auditory stimuli influences the perceived visual flicker rate. The influence of auditory rate is particularly strong when it is faster than the visual rate and temporal after-effects analogous to the ventriloquism after-effect have also been reported (Recanzone, 2003). A simplification of the auditory temporal driving of visual perception was recently demonstrated in which the rapid presentation of two transient sounds paired with a single flash lead to the perception of two flashes (Shams et al., 2000; Fig. 11.6C). This phenomenon is robust and likely entails low-level interactions between the auditory and visual sensory modalities as it resists intense perceptual training designed to alleviate the illusory report (Rosenthal et al., 2009). A few studies have started to address these effects with neuroimaging and have shown correlates of the illusions emerging in visual cortices (Shams et al., 2005; Watkins et al., 2006; Noesselt et al., 2008) without specific auditory cortex modulation in this context.

In a design that did not specifically test for illusory effects, Noesselt et al. (2007) used temporally coincident or noncoincident AV events that were presented in a visual brightness change detection task. This fMRI study was designed to evaluate the impact of AV temporal coincidence for meaningless events. The authors reported several regions sensitive to the temporal coincidence of AV events that were all contralateral to the side of stimuli presentation: the middle STS (mSTS), the visual, and the auditory cortices. The authors specifically reported an increased activation of the medial part of Heschl's gyrus (HG) during coincident AV presentations, and this activation extended to the posterior insula and the planum temporale (PT). The authors suggested that the increased activation in A1 during coincident AV presentation resulted from a possible feedback from mSTS.

Modulations of auditory and visual sensory evoked responses have often been reported during presentation of coincident AV stimuli (Giard & Peronnet, 1999; Foxe et al., 2000; Molholm et al., 2002). The presentation of coincident nonspeech AV events can increase sensory evoked responses, with modulations observed as early as 50 ms; the amplitude modulations are function of task (Fort & Giard, 2004) but can also be observed in the absence of a specific task (Vidal et al., 2008). Martuzzi and colleagues (2007) showed with fMRI that activity in primary auditory and visual sensory cortices was sensitive to sensory information presented in the other modality and to coincident AV presentation. The first fMRI report of auditory cortex activation in the presence of visual stimuli was provided by Calvert et al. (1997), who showed activation in primary and secondary auditory cortices (BA 41, 42, and 22) in response to the presentation of a speaking face. The specific implication of auditory cortex in the integration of AV speech is addressed in Section 11.4.

The modulatory effects on auditory cortical responses in the context of multisensory perception converge with the neurophysiological findings described in Section 11.2. Further convergent evidence specifically aiming at bridging findings in monkeys and humans has been provided by Kayser et al. (2007), who recorded anesthetized and awake macaque monkeys with fMRI while they were presented with auditory, visual, and AV natural scenes. To allow proper functional mapping, auditory cortex was first functionally parcellated using bandwidth and frequency mapping, leading to distinct belt and parabelt fields and core auditory region. In anesthetized monkeys, the authors reported activation of caudal auditory cortex in response to visual stimuli alone and enhanced activation throughout auditory cortex in response to AV stimuli compared to audio-alone stimulation. Specifically, the caudal (CM) and caudolateral (CL) fields showed significant responses to visual stimulation alone; AV presentations lead to enhanced responses in the CL and mediomedial (MM) belt fields and in primary auditory cortex. AV presentations activated several regions in the caudal and lateral region of auditory cortex, and more consistently so in the CM, CL, and MM fields. In alert behaving monkeys, identical visual stimuli elicited significant activation in CL, CM, MM, rostromedial field, and primary auditory cortex (R, A1). The presentation of AV stimuli enhanced activation in CL, CM, and A1. In both anesthetized and awake monkeys, visual and AV stimulation activated the caudal more than the rostral parabelt. Kayser et al. (2008) demonstrated that visual modulation of auditory cortex could be found both in field potentials and in single-unit firing rate. Visual modulation of auditory cortex activity was strongest when visual stimuli preceded by 20–80 ms the auditory onset. This finding further corroborates the latency patterns of sensory responses outlined in Section 11.2: specifically, AV events show natural asynchronies in favor of visual events preceding auditory events, suggesting potential evolutionary adaptation of sensory processing times. In other words, modulation of auditory cortex response by visual inputs may entail precedence of visual inputs over auditory ones. This observation is crucial for ecologically relevant stimuli such as speech (Section 11.4).

Few studies have specifically addressed the effect of desynchronized AV signals on cortical responses (Bushara et al., 2001; Miller & D’Esposito, 2005; Doesburg et al., 2008). fMRI findings suggest that two networks can be distinguished on the basis of the sensitivity to (1) the coincidence of multisensory stimuli and (2) the fused or unfused perceptual outcome (Miller & D’Esposito, 2005). In this study, regions that showed sensitivity to desynchronized stimuli regardless of the perceptual outcome comprised the superior colliculus, the anterior insula, and the anterior inferior parietal sulcus (IPS) whereas regions sensitive to the perceptual outcome independently of the asynchrony of the stimuli were primary auditory cortex (HG), mSTS, mIPS, and Inferior Frontal Gyrus (IFG) (Miller & D’Esposito, 2005). Bushara and colleagues (2001) assigned a particular role to the insula when participants were presented with desynchronized nonspeech AV stimuli. An additional fMRI study (Calvert et al., 2001) tested which brain regions were sensitive to temporal (de)synchrony of a visual checkerboard and an auditory noise burst. The stimuli were presented at a rate of 8 Hz and could either be in synchrony or randomly desynchronized. The most sensitive regions to temporal properties was the superior colliculus along with the left STS which showed supra- and sub-additive responses to synchronized and desynchronized periods of AV stimulation, respectively. No auditory cortex activation was reported.

Cumulatively, the results suggest that the *functional contribution of auditory (and visual) cortices to the analysis of multisensory events is specific to the perceptual outcome and not reducible to the spatiotemporal coincidence of the stimuli*. Whether information presented in the visual modality modulates activation of auditory cortex in a direct manner or is relayed via multisensory regions (e.g., mSTS) remains unclear.

Additional complications emerge in complex perceptual tasks during which attentional factors cannot be set aside. For instance, the salience of an event in one sensory modality can be increased by the concomitant presentation of an event in a different sensory modality (Stein et al., 1996; McDonald & Ward, 2000). It is well known that attention can significantly modulate the response properties of auditory cortical responses (e.g., Jancke et al., 1999). Conversely, in multisensory contexts, paying attention to one modality (and disregarding the other for task purposes) leads to depressed activation in the cortices of the nonattended sensory modality: hence, being engaged in visual tasks has been shown to depress responses in auditory cortices (Haxby et al., 1994) and similarly, depress responses in visual cortices are observed when engaged in somatosensory tasks (Kawashima et al., 1995). The issue of separate versus shared attentional mechanisms in multisensory perception are ongoing and drive to a great extent the interpretation of early modulations of auditory cortex activation.

11.4 Multisensory Speech and Language

The functional implication of auditory cortex in the integration of auditory and visual (and to some extent tactile) information in speech perception has been a particularly productive area of investigation. Some key findings are summarized to provide a general context for AV speech integration and to highlight the level of specificity in interpreting activation of auditory cortex in such contexts.

11.4.1 Classic Findings in AV Speech

11.4.1.1 Contribution of Visual Speech to Auditory Speech

It is well known that seeing a person's face influences the perception of auditory speech (Summerfield, 1987; Green, 1998; Campbell, 2008). The influence of visual speechreading in the comprehension of auditory speech was quantified in the 1950s by Sumby and Pollack (1954). In their seminal study, participants watched a speaker articulating words and listened to the corresponding auditory utterances presented with variable signal-to-noise ratios (SNRs). The authors found that the benefit of seeing the face in comprehending auditory words varied as a function of SNR and size of vocabulary: the contribution of speechreading to auditory word recognition increased as the SNR and vocabulary size diminished (as tested, down to eight words); at higher

SNR, the contribution of visual speech was mostly seen for large vocabulary sizes (as tested, up to 256 words). These findings and others (Erber, 1969; Binnie et al., 1974) suggested that visual speech contributes to auditory speech comprehension mainly under adverse listening conditions, an analogue to improving SNR. Recent investigation has suggested that maximal integration in AV speech was in fact observed within a medium range of auditory SNR with a maximal benefit at –12 dB (Ross et al., 2007). A major cue underlying the integration of AV speech information is the correlation between the envelope of the acoustic signal and the visible movements of the facial articulators (Grant & Seitz, 2000): specifically, major AV benefits are obtained using low-frequency rather than high-frequency regions of the auditory spectrum (Grant & Walden, 1996). A gain in auditory comprehension is also obtained when using hard to comprehend sentences (Reisberg et al., 1987), suggesting that visual information does not solely benefit noisy acoustic signals but also the linguistic content of auditory speech. Additional findings suggest that lexical context can constrain AV speech integration (Brancazio, 2004; Barutchu et al., 2008). Importantly, it has been suggested that the perceptual processes underlying AV speech integration in sentences and nonsense syllables are likely to differ, as measures of integration in one condition do not necessarily predict benefits in the other (Grant & Seitz, 1998).

11.4.1.2 McGurk and MacDonald Effects

In their classic study, McGurk and MacDonald (1976) showed novel robust illusions in which visual information influences the phonetic categorization of auditory speech. Two kinds of illusions were reported: *fusion* and *combination*. The term "McGurk illusion" has often been used to refer to the fusion case. In the fusion case, a visual velar (e.g., "ga" or "ka") simultaneously presented with an auditory bilabial (e.g., "ba" or "pa") leads to a fused and unique alveolar/dental percept (e.g., "tha" or "ta," respectively). In the combination case, presenting an auditory velar dubbed onto a face articulating a bilabial leads to multiple perceptual outcomes or combinations such as "pka," "kapa," etc. McGurk illusions have served as direct markers of AV speech integration. The McGurk fusion effect has been extensively studied and often considered to be an automatic case of AV integration, that is, an example of multisensory integration independent of attention (Soto-Faraco et al., 2004). Often, despite being told about the illusion and the nature of stimuli being displayed, participants will still report a robust illusory percept (Summerfield & McGrath, 1984; Rosenblum & Saldãna, 1996). However, recent studies have demonstrated that AV speech integration may not be entirely immune to attentional effects (Tiippana et al., 2004; Alsius et al., 2005; Munhall et al., 2009).

11.4.1.3 Constraints on AV Speech Integration

A classic approach in determining the constraints on the integration of multisensory events is to impoverish the incoming signals by (1) introducing noise in the sensory modalities under study – hence, testing the inverse effectiveness principle or (2) misaligning the events in space or desynchronizing the events in time, hence testing the spatiotemporal coincidence principle.

Spatial Resolution. A straightforward approach to determining the spatial resolution necessary for visual speech information is to vary the viewing distance. The earliest study addressing the influence of distance on visual speech perception was conducted by Erber (1971), who tested distances ranging from 5 to 100 feet (~1.5–30 m) in deaf children: their performance in lipreading dropped from 75% to 11% with increased distance. Measures of speechreading ability are a good predictor of AV speech integration performance (Grant et al., 1998) and distance should thus affect AV speech performance. Jordan and Sergeant (2000) used a similar set of distances (1–30 m) to evaluate AV speech perception of congruent and incongruent (McGurk) speech: the presence of visual speech outperformed auditory speech recognition alone at all distances. The relative independence of AV speech integration with visual distance thus suggested that low spatial frequencies contribute most to speechreading ability. In a first attempt at quantifying the contribution of visual speech information to auditory comprehension as a function of visual spatial frequencies, Erber (1979) showed that visual benefits reached a plateau under increased blurring of the face. Recent findings show that even noisy visual information can robustly influence auditory categorization; for instance, McGurk fusion is resilient to the filtering out of facial information (Campbell & Massaro, 1997; MacDonald et al., 2000) and can be reliably obtained when replacing a face by point-of-lights display (Rosenblum & Saldaña, 1996). Munhall et al. (2004) evaluated the effect of bandpass and low-pass spatial filtering (2.7–44.1 cycles/face) of the face in noisy listening conditions. They reported an enhanced auditory intelligibility in all filtering conditions with a maximal enhancement for approximately11 cycles/face. Low-passed filtered facial displays showed equivalent results as nonfiltered displays, suggesting that the low spatial frequencies carry the relevant information for the speech system.

Temporal Resolution. A second approach in determining the constraints on AV speech integration is to quantify the temporal tolerance of AV speech to desynchronized inputs. Nonspeech results have shown that the optimal timing for nonspeech AV integration is a visual lead of approximately 60 ms (see Section 11.3). In one of the earliest studies on AV speech tolerance to desynchrony, Pandey et al. (1986) found that an audio delay of up to 80 ms did not affect the visual benefits of seeing the speaker’s face. They interpreted these data as evidence that audio delays in connected speech are disruptive at the syllabic level but not at the phonemic recognition stage. Dixon and Spitz (1980) compared a participant’s sensitivity to desynchronized connected speech and to a video of a hammer hitting a peg. In both cases, the tolerance to AV desynchrony was much higher than in nonnatural events: participants tolerated nearly 260 ms and 190 ms of auditory delay and 130 ms and 75 ms of auditory leads in speech and hammer contexts, respectively. For speech-like stimuli, 80–140 ms has been estimated to be the best temporal resolution human observers can have; AV desynchronies below 80 ms do not disturb comprehension of connected speech (McGrath & Summerfield, 1985). These results suggest that precise temporal synchrony in AV speech may not be necessary for AV speech integration.

Studies specifically addressing the tolerance of AV speech integration reveal that optimal integration spans a range of nearly 300 ms for nonsense syllables (Munhall et al., 1996; Conrey & Pisoni, 2006; van Wassenhove et al., 2007). In all these studies, an asymmetry between auditory and visual leads was observed, whereby visual leads were better tolerated than auditory leads. The functional implication of this finding is in line with the natural desynchrony of AV events: a recent quantification using French and English AV speech database has shown that visual speech information often precedes by as much as 100–300 ms the auditory information (Chandrasekaran et al., 2009).

AV Speech Perception in Development. By 2 months of age, infants can match articulating faces with their corresponding auditory utterances (Dodd, 1979; Kuhl & Meltzoff, 1982; Patterson & Werker, 2003) and by 5 months of age, infants show signs of McGurk-like effects (Burnham & Dodd, 2004; Rosenblum et al., 1997). A recent study on AV speech matching in 9- to 12-week-old infants showed in two EEG experiments that the early auditory responses were sensitive to visual speech inputs at the phonetic level (Bristow et al., 2008). In this study, a mismatch response was observed when visual speech was incongruent with auditory speech and cortical sources involved in the mismatch responses were comparable to those observed in adults, namely composed of the left STS, STG, supramarginal gyrus (SMG), and IFG. In this network, a pattern of repetition suppression was observed in temporal regions whereas repetition enhancement was observed in frontal regions for incongruent AV stimuli. The authors suggested that this pattern could reflect processes by which infants learn and store AV speech templates later on used for matching their speech production (Bristow et al., 2008). Early signs of AV speech integration are in line with additional findings showing the existence of a critical period for AV speech integration at about 2 years of age (Schorr et al., 2005).

The evolution of speechreading abilities over the course of development and its implication in the efficiency of AV speech integration needs further investigation. In particular, the STG and angular gyrus are already active in speech processing at 3 months of age (Dehaene-Lambertz et al., 2002), and the earlier and faster development of the auditory system compared to the visual system during development could account for auditory dominance over speechreading (Dubois et al., 2008).

11.4.2 AV Speech Binding

A majority of speech perception models have overlooked speechreading as a possible source of linguistic information leading to the problem of *when* visual information (*where* and *how* in the brain) integrates with auditory-based speech processing. In "early models," auditory and visual features are integrated before phonologic categorization whereas in "late models," it is after auditory speech has been categorized that visual information influences the representational outcome. Several dichotomies for AV speech processing models have been described in detail (Summerfield, 1987; Schwartz et al., 1998). In the *direct identification model*, AV speech coincides with the decision stage, implying that sensory-specific information is in a common readable format at the integration stage. For instance, the pre-labeling model proposed by Braida (1991) assumed a common representational metric for AV speech processing. In the *dominant recoding model*, visual information is recoded in an auditory form (the dominant form in speech perception) before being integrated with the incoming auditory information. In an analogous recoding strategy, the underlying articulatory gestures of speech could be the metric by which auditory speech is being processed for instance in the *motor theory of speech perception* proposed by Liberman et al. (Liberman et al., 1967: Liberman & Mattingly, 1985). By extension, visual speech may follow the same encoding procedure. On arrival to the integration stage, auditory and visual speech information are thus in a motoric form. A prominent computational approach to the problem of AV speech integration is a *separate identification model* essentially based on the *fuzzy-logical model of perception* (FLMP) proposed by Massaro (1987). The initial proposal was that auditory and visual speech inputs were independently evaluated before being integrated thus accounting for a late integration model. More recently, a second locus of interaction—evaluation stage—has been added before the integration, now a decision stage (Massaro, 1998). An alternative model based on neuroimaging data has recently been proposed (van Wassenhove et al., 2005; Poeppel et al., 2008) on the basis of the *analysis-by-synthesis* model of speech perception (Halle & Stevens, 1967). In this model, visual speech predicts the incoming auditory information owing to the inherent precedence of articulatory facial movements (Chandrasekaran et al., 2009). Recent findings provide further evidence for comparable predictive models in AV speech processing implicating unisensory and multisensory regions (e.g., Arnal et al., 2009, 2011).

11.4.3 Functional Brain Imaging Findings

AV speech processing research has greatly benefited from advances in neuroimaging and a few results have now been replicated across diverse imaging techniques.

11.4.3.1 Contribution of Auditory Cortex to Visual Speechreading

The first evidence for the implication of primary and secondary auditory cortices in visual speech processing in the absence of any auditory inputs was demonstrated using fMRI (Calvert et al., 1997, 2001). Supra-additivity in primary and secondary auditory cortices was tested as an index of multisensory integration and taken as evidence for multisensory integration in the vicinity of primary auditory cortex (Calvert et al., 2001). Additional fMRI studies have suggested that primary auditory cortex (PAC) is implicated in the processing of visual speech (Pekkola et al., 2005) with reports of bilateral and left-lateralized activations of auditory cortex to the presentation of visual speech alone (Capek et al., 2004). The activation of auditory cortex was found in response to the presentation of articulating faces and during silent lipreading (MacSweeney et al., 2000; Sato et al., 2004) but not to the presentation of still faces (Calvert & Campbell, 2003). Activation of PAC to the presentation of visual speech has thus been argued to be specific to the linguistic content provided by the speaking face (Calvert, 1997; Campbell et al., 2008 Pekkola et al., 2005).

The activation of PAC in response to visual speech still remains controversial. For instance, in a study using synthetic visual speech, Wright and colleagues (2003) reported activation in the right posterior STS to the presentation of mouthed words located about *50 mm away from* but *not in* auditory cortex. In an attempt to determine whether PAC was activated by visual only stimuli, Bernstein and colleagues (2002) used auditory localizers to map out the temporal plane and PAC in each participant. They then presented participants with a series of still faces, non speech stimuli and visual speech mouthing monosyllabic words. Lipreading was found to activate the superior temporal plane, the STS, and the middle temporal gyrus (MTG) as well as several frontal areas (inferior, middle, and superior frontal gyri); however, the only activated region of overlap in the auditory localizer and in the lipreading conditions was located along the lateral surface of the STG. Thus, in this study (Bernstein et al., 2002) and others (Olson et al., 2002), no evidence for PAC activation was found in the lipreading condition alone. Paulesu and colleagues (2003) used positron emission tomography (PET) to contrast regions responsive to visual mouthing of bisyllabic high-frequency words (lexically identifiable), the same words played backwards (not lexically identifiable) and still faces. Neither types of stimuli lead to PAC activation but activation of secondary associations areas were reported for words.

Whether PAC is engaged in processing visual speech remains highly controversial, and substantial fMRI and PET findings for or against it are available. Two confounds in using fMRI and PET methodologies are (1) the careful mapping of auditory cortex needs to be conducted before establishing which zone of auditory cortex is indeed activated by visual lipreading alone (a technique often used in vision research) and (2) the response profiles in this area are likely to be diverse and subject to the type of analyses being used for quantification. In fact, a recent study using intracranial recordings in humans found 13 response profiles in the temporal lobe alone after the presentation of visual speech (Besle et al., 2008). Besle and colleagues reported early responses to visual speech inputs (preceding by 100 ms the auditory onset for AV material) in the posterior MTG and superior temporal cortex including the HG, the planum temporale, the planum porale, the STG, and the STS. Overall, this study provides robust evidence in humans for responses of the secondary but not primary auditory cortices to the presentation of visual speech.

11.4.3.2 Contribution of Auditory Cortex to the Perception of AV Speech

The first brain imaging study focusing on AV speech perception was conducted by Sams et al. (1991), who used an MMN paradigm with MEG to evaluate whether sources of the auditory magnetic responses were sensitive to visual speech information.

The authors used congruent and incongruent (McGurk: audio /pa/ dubbed onto visual /ka/) stimuli. They found that the presentation of an incongruent (congruent) AV deviant in a stream of congruent (incongruent) AV standards elicited a robust auditory MMN. The reconstruction of the equivalent current dipole (ECD) showed that the sources prior to and at the origin of the recorded MMN were located in the left supratemporal plane. The authors suggested that the generators were likely to be located in the PAC and surrounding auditory belt areas. A series of MMN studies in AV speech has since replicated these findings (Colin et al., 2002, 2004; Möttönen et al., 2002, 2004). Sources obtained with the MMN paradigm have been located in auditory association areas at about 150–200 ms following auditory onset and in the STS from 250 ms on. Consistent with MMN studies of AV speech, Klucharev and colleagues (2003) used auditory, visual, and AV vowels (congruent and incongruent) for their EEG study. They reported several times at which multisensory interactions occurred: the early stage showed no linguistic specificity (before ~150 ms) and a subsequent stage (after ~150 ms) was reported to be sensitive to phonetic content. The sources accounting for the modulation of the auditory responses at those timings were located in auditory association cortices.

Evidence for the modulation of auditory cortex by visual speech inputs was also observed in an MEG study (Jääskeläinen et al., 2004) in which participants were shown a video of a face articulating the same or a different vowel sound that was displayed 500 ms after the presentation of the face: the amplitude of the auditory evoked responses was decreased under such conditions, suggesting that visual speech inputs lead to adaptation of the subset of auditory neurons responsive to that feature. However, no difference in amplitude was observed when the visual stimuli were drawn from a same or from a different phonetic category. Several EEG studies of AV speech processing (Fig. 11.7) have similarly reported a suppressed auditory evoked response to the presentation of synchronized AV speech compared to auditory speech (Besle et al., 2004; van Wassenhove et al., 2005; Pilling, 2009; Arnal et al., 2009). Two studies reported that the latency but not the amplitude reduction of the auditory evoked response was a function of the phonetic content provided in visual speech (van Wassenhove et al., 2005; Arnal et al., 2009; Fig. 11.7B–E). The amplitude reduction of the auditory evoked responses observed in EEG and MEG to the presentation of AV speech is supported by intracranial recordings (Reale et al., 2007; Besle et al., 2008). In particular, Besle and colleagues (2008) reported two kinds of AV interaction in the secondary auditory association cortices after the first influence of visual speech in this region. At the onset of the auditory syllable, the initial visual influence disappears and the amplitude of the auditory response is decreased compared to auditory alone presentation. Similar amplitude reductions were observed to the presentation of AV syllables over the left lateral pSTG (Reale et al., 2007).

It is noteworthy that the MEG, EEG, and surface EEG (sEEG) reports sharply contrast with fMRI and PET findings in which enhanced and supra-additive BOLD activations to the presentation of visual and AV speech are reported in the middle and posterior STG and posterior STS (Calvert et al., 1997; Skipper et al., 2007). An increased (Calvert et al., 1999) to supra-additive (Calvert et al., 2000) activation

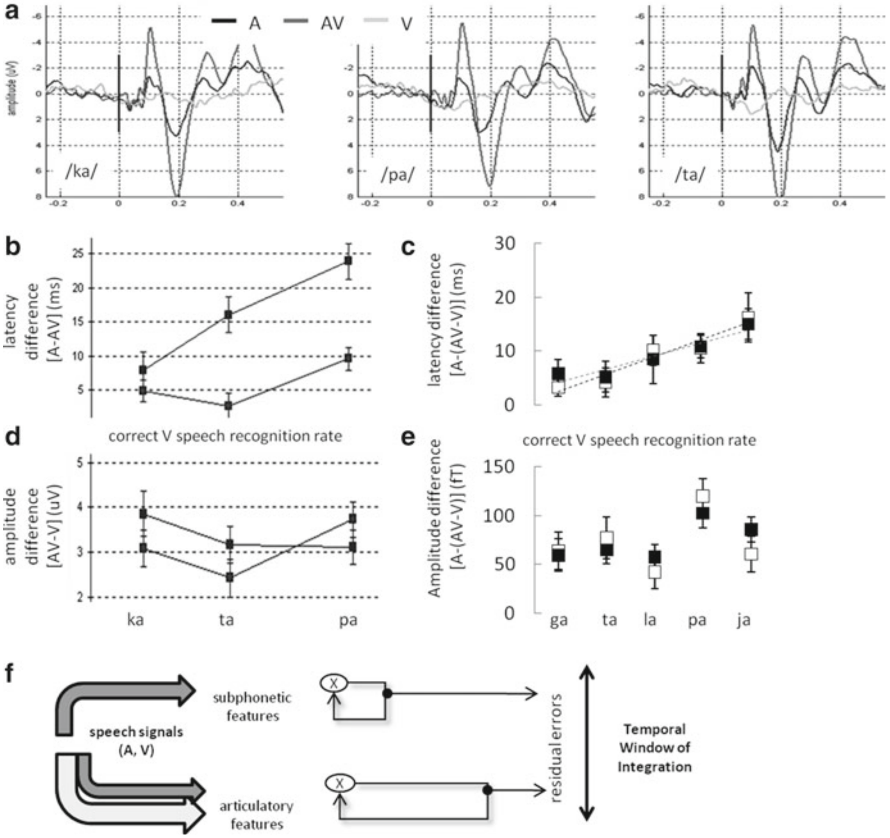


Fig. 11.7 AV speech processing. (**a**) Auditory-evoked responses recorded at a frontocentral electrode (EEG) to the presentation of auditory (black), visual (light gray), and audiovisual (dark gray) syllables /ka/, /pa/, and /ta/ (van Wassenhove et al., 2005). No visible auditory evoked-responses were observed for visual alone speech syllables; as can be readily observed, auditory evoked-responses to the presentation of auditory speech alone were found to be larger in amplitude than to the presentation of audiovisual speech. (**b–e**) Two independent studies (van Wassenhove et al., 2005; Arnal et al., 2010) demonstrated that the saliency in visual speech (derived from participants' correct recognition rate of visual speech) is indicative of the extent to which the auditory evoked responses will be shifted in time (**b, c**) but not to which its amplitude will be decreased (**d, e**). The more informative the visual speech event, the more predictable the auditory target and the faster the auditory response (**b, c**). The amplitude decrease of early auditory evoked responses remains independent of visual speech saliency; rather, it may index the perceived incongruence of audiovisual speech (cf. Arnal et al., 2010). (**f**) from these results, an analysis-by-synthesis model for audiovisual speech processing was proposed in which the natural precedence of visual speech predicts the auditory targets (van Wassenhove et al., 2005). The analysis of speech on two different resolutions were posited and their integration taken as reflecting the temporal window of integration classically observed in audiovisual speech integration (see text). More details on analysis-by-synthesis in speech processing are provided in another chapter (Giraud and Poeppel, Chapter 9)

has been reported in the STS and STG and sub-additivity in these same regions together with left inferior temporal gyrus (BA 44/45), premotor cortex (BA 6), and anterior cingulate gyrus (BA 32) to the presentation of congruent and incongruent AV speech, respectively. In their EEG study, Callan et al. (2001) found a greater enhancement in the high-frequency band (45–70 Hz) for AV speech stimuli with audio noise compared to auditory stimuli alone. CSD analysis revealed a main source located in the STG. However, using an MMN design in which standards were congruent and deviants were McGurked AV syllables, Kaiser and colleagues (2005) found no significant differences in oscillatory activity over auditory cortices for incongruent tokens; enhanced high gamma band (~75 Hz) responses were found for incongruent oddballs on sensors located over parietal cortex, occipital cortices, and left inferior frontal cortices. The authors suggested that activity over auditory cortex may have been too transient for their analysis to capture significant effects.

Other fMRI findings (Callan et al., 2003) showed significant activation of the MTG, STS, and STG in response to the presentation of AV speech in noise; BOLD activation consistent with the inverse effectiveness principle in these same regions (MTG, STS, and STG) was reported for stimuli providing information on the place of articulation (Callan et al., 2004). The left posterior STS has also been shown to be sensitive to incongruence in AV speech (Calvert et al., 2000; Wright et al., 2003; Miller & D'Esposito, 2005). Using fMRI and PET, Sekiyama and colleagues (2003) used the McGurk effect with two levels of auditory noise; comparison between the low and high SNR conditions revealed a left lateralized activation in the posterior STS and BA 22, thalamus, and cerebellum. However, not all studies support the inverse effectiveness principle in auditory cortex (Calvert et al., 1999; Jones & Callan, 2003; Sadato et al., 2004). Desynchronizing AV McGurk syllables does not significantly affect activation of the STS or auditory cortex (Olson et al., 2002; Jones & Callan, 2003) whereas others report significant and systematic activation of HG as a function of desynchrony (Miller & D'Esposito, 2005).

Overall, it has become clear that auditory cortex is modulated by visual speech inputs in the context of AV speech processing but the specificity of this modulation remains unsettled. Several studies have suggested that these modulations were specific to perceptual (phonetic) outcome such that visual speech information acts as a predictor of incoming auditory targets (van Wassenhove et al., 2005; Skipper et al., 2007; Arnal et al., 2009, 2011); thereby auditory speech processing consists in computing the residual error between incoming visual speech information and the auditory target (Fig. 11.7F). Supporting evidence in human research for this interpretation is of two kinds: facial kinematics naturally precede auditory speech (Chandrasekaran et al., 2009) and intracranial recordings indicate that visual speech can modulate the responsiveness of auditory cortex early on (Besle et al., 2008). Based on a set of neurophysiological recordings in monkeys, it has been proposed that visual inputs (which predictably lead auditory inputs) can change the excitability of auditory cortex by resetting the phase of ongoing oscillation (Schroeder et al., 2008), thereby amplifying auditory cortex responses. Recent MEG findings suggest that this mechanism is obtained in binding AV information in humans as well (Luo et al., 2010).

11.4.3.3 Contribution of Auditory Cortex to Reading

After the effortful learning of matching letters and sound attributes, reading usually becomes an effortless skill in literate adults; the association "sound-letter" is a prime example of extensive learning during development. In a first MEG study addressing the issue of multisensory integration during reading, Raij and colleagues (2000) used auditory, visual, and AV presentations of matching and nonmatching letters and unpronounceable controls. The first evidence of AV integration was observed approximately 200 ms after sensory-specific responses. Source reconstructions showed that activation to matching letters was stronger than to non matching letters as demonstrated by a clear bilateral differences in STS starting approximately 400 ms poststimulus onset (which is rather late, given the processes recruited). No activation or modulation of auditory cortices was shown in this study. However, using a similar paradigm in fMRI, van Atteveldt and colleagues (2004) showed that the activation of primary auditory cortex was modulated by the presentation of letters with a stronger activation for congruent than for incongruent letter–sound pairing. Activation to incongruent letter–sound pairing in the PT and HG was significantly smaller than activation to congruent letter–sound pairs and to sound alone. An EEG study using an MMN paradigm in which an enhanced mismatch to incongruent letter–sound pairing progressively decreased with stimuli desynchrony (up to 200 ms) further suggests an early and automatic implication of auditory association cortex in the integration of sound–letter information (Froyen et al., 2008). Over the few studies addressing the topic of sound–letter matching, a dichotomy between suppression effects recorded with MEG and enhanced activity recorded with fMRI is reminiscent of the findings in AV speech integration. These two imaging techniques provide complementary anatomical and temporal resolution and the data suggest that both top-down and feedforward connectivity to auditory cortex is implicated in the mapping of auditory and visual letters to phonological code (van Atteveldt et al., 2009, 2010).

11.5 Summary

In some 30 years of multisensory research, it has become evident that sensory areas once considered as independent dedicated pathways for the analysis of specific features in one sensory modality are in fact contributing to a diverse array of computations, including those a priori pertaining to other sensory modalities.

This chapter aimed at illustrating the current neurophysiological understanding of auditory cortex with respect to its modulation *by* and *of* other sensory areas. The diversity of empirical observations implicating the auditory areas in multisensory perception highlights that their contribution is far from fully understood. For instance, the more general modulatory effects readily observed in neurophysiological and neuroimaging studies stand in sharp contrast with the perceptual specificities described throughout the chapter. As such, much research is still needed to specify to what extent the contribution of auditory cortices is at times sufficient and at others necessary for particular sensory analyses and computations in multisensory perception.

This chapter further aimed at illustrating the need to define to what extent a sensory area is indeed functionally dedicated to the analysis of particular sensory inputs and as such, highlight the important contribution of multisensory research in determining the roots of functional specialization in cortex.

References

Alais, D., & Burr, D. (2003). The "flash-lag" effect occurs in audition and cross-modally. *Current Biology*, 13(1), 59–63.

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology*, 15(9), 839–843.

Arnal, L. H., Morillon, B., Kell, C. A., & Giraud, A.-L. (2009). Dual neural routing of visual facilitation in speech processing. *Journal of Neuroscience*, 29(43), 13445–13453.

Arnal, L. H., Wyart, V., & Giraud, A.-L. (2011). Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nature Neuroscience*, 14(6), 797–801.

Barutchu, A., Crewther, S. G., Kiely, P., Murphy, M. J., & Crewther, D. P. (2008). When /b/ill with /g/ill becomes /d/ill: Evidence for a lexical effect in audiovisual speech perception. *European Journal of Cognitive Psychology*, 20(1), 1–11.

Benevento, L. A., Fallon, J., Davis, B. J., & Rezak, M. (1977). Auditory-visual interaction in single cells in the cortex of the superior temporal sulcus and the orbital frontal cortex of the macaque monkey. *Experimental Neurology*, 57(3), 849–872.

Bernstein, L. E., Auer, E. T. J., Moore, J. K., Ponton, C. W., Don, M., & Singh, M. (2002). Visual speech perception without primary auditory cortex activation. *NeuroReport*, 13(3), 311–315.

Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Attention*, *Perception*, *& Psychophysics*, 29(6), 578–584.

Bertelson, P., Vroomen, J., De Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Attention*, *Perception*, *& Psychophysics*, 62(2), 321–332.

Besle, J., Fort, A., Delpuech, C., & Giard, M.-H. (2004). Bimodal speech: Early suppressive visual effects in human auditory cortex. *European Journal of Neuroscience*, 20(8), 2225–2234.

Besle, J., Fischer, C., Bidet-Caulet, A., Lecaignard, F., Bertrand, O., & Giard, M.-H. (2008). Visual activation and audiovisual interactions in the auditory cortex during speech perception: Intracranial recordings in humans. *Journal of Neuroscience*, 28(52), 14301–14310.

Besle, J., Bertrand, O., & Giard, M.-H. (2009). Electrophysiological (EEG, sEEG, MEG) evidence for multiple audiovisual interactions in the human auditory cortex. *Hearing Research*, 258(1–2), 143–151.

Binnie, C. A., Montgomery, A. A., & Jackson, P. L. (1974). Auditory and visual contributions to the perception of consonants. *Journal of Speech and Hearing Research*, 17(4), 619–630.

Bischoff, M., Walter, B., Blecker, C. R., Morgen, K., Vaitl, D., & Sammer, G. (2007). Utilizing the ventriloquism-effect to investigate audio-visual binding. *Neuropsychologia*, 45(3), 578–586.

Bizley, J. K., Nodal, F. R., Bajo, V. M., Nelken, I., & King, A. J. (2007). Physiological and anatomical evidence for multisensory interactions in auditory cortex. *Cerebral Cortex*, 17(9), 2172–2189.

Bonath, B., Noesselt, T., Martinez, A., Mishra, J., Schwiecker, K., Heinze, H.-J., & Hillyard, S. A. (2007). Neural basis of the ventriloquist illusion. *Current Biology*, 17(19), 1697–1703.

Braida, L. D. (1991). Crossmodal integration in the identification of consonant segments. *The Quarterly Journal of Experimental Psychology Section A*, 43(3), 647–677.

Brancazio, L. (2004). Lexical influences in audiovisual speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 30(3), 445–463.

Bristow, D., Dehaene-Lambertz, G., Mattout, J., Soares, C., Gliga, T., Baillet, S., & Mangin, J.-F. (2008). Hearing faces: How the infant brain matches the face it sees with the speech it hears. *Journal of Cognitive Neuroscience*, 21(5), 905–921.

Brosch, M., Selezneva, E., & Scheich, H. (2005). Nonauditory events of a behavioral procedure activate auditory cortex of highly trained monkeys. *Journal of Neuroscience*, 25(29), 6797–6806.

Budinger, E., Heil, P., Hess, A., & Scheich, H. (2006). Multisensory processing via early cortical stages: Connections of the primary auditory cortical field with other sensory systems. *Neuroscience*, 143(4), 1065–1083.

Burnham, D., & Dodd, B. (2004). Auditory–visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology*, 45(4), 204–220.

Burton, H., & Sinclair, R. J. (1996). Somatosensory cortex and tactile perceptions. In L. Kruger (Ed.), *Pain and touch* (pp. 105–177). San Diego, CA.

Bushara, K. O., Grafman, J., & Hallett, M. (2001). Neural correlates of auditory–visual stimulus onset asynchrony detection. *Journal of Neuroscience*, 21(1), 300–304.

Caclin, A., Soto-Faraco, S., Kingstone, A., & Spence, C. (2002). Tactile "capture" of audition. *Attention*, *Perception*, *& Psychophysics*, 64(4), 616–630.

Caetano, G., & Jousmäki, V. (2006). Evidence of vibrotactile input to human auditory cortex. *NeuroImage*, 29(1), 15–28.

Callan, D. E., Callan, A. M., Kroos, C., & Vatikiotis-Bateson, E. (2001). Multimodal contribution to speech perception revealed by independent component analysis: A single-sweep EEG case study. *Cognitive Brain Research*, 10(3), 349–353.

Callan, D. E., Jones, J. A., Munhall, K., Callan, A. M., Kroos, C., & Vatikiotis-Bateson, E. (2003). Neural processes underlying perceptual enhancement by visual speech gestures. *NeuroReport*, 114(17), 2213–2218.

Callan, D. E., Jones, J. A., Munhall, K., Kroos, C., Callan, A. M., & Vatikiotis-Bateson, E. (2004). Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information. *Journal of Cognitive Neuroscience*, 16(5), 805–816.

Calvert, G. A., & Campbell, R. (2003). Reading speech from still and moving faces: The neural substrates of visible speech. *Journal of Cognitive Neuroscience*, 15(1), 57–70.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P. K., et al. (1997). Activation of auditory cortex during silent lipreading. *Science*, 276(5312), 593–596.

Calvert, G. A., Brammer, M. J., Bullmore, E. T., Campbell, R., Iversen, S. D., & David, A. S. (1999). Response amplification in sensory-specific cortices during crossmodal binding. *NeuroReport*, 10(12), 2619–2623.

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology*, 10(11), 649–657.

Calvert, G. A., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *NeuroImage*, 14(2), 427–438.

Campbell, C. S., & Massaro, D. W. (1997). Perception of visible speech: Influence of spatial quantization. *Perception*, 26(5), 627–644.

Campbell, R. (2008). The processing of audio-visual speech: Empirical and neural bases. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493), 1001–1010.

Canon, L. K. (1970). Intermodality inconsistency of input and directed attention as determinants of the nature of adaptation. *Journal of Experimental Psychology*, 84(1), 141–147.

Capek, C. M., Bavelier, D., Corina, D., Newman, A. J., Jezzard, P., & Neville, H. J. (2004). The cortical organization of audio-visual sentence comprehension: An fMRI study at 4 Tesla. *Cognitive Brain Research*, 20(2), 111–119.

Cappe, C., Morel, A., Barone, P., & Rouiller, E. M. (2009). The thalamocortical projection systems in primate: An anatomical support for multisensory and sensorimotor interplay. *Cerebral Cortex*, 19(9), 2025–2037.

Celesia, G. G. (1968). Auditory evoked responses: Intracranial and extracranial average evoked responses. *Archives of Neurology*, 19(4), 430–437.

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS Computational Biology*, 5(7), e1000436.

Chen, C.-M., Lakatos, P., Shah, A. S., Mehta, A. D., Givre, S. J., Javitt, D. C., & Schroeder, C. E. (2007). Functional anatomy and interaction of fast and slow visual pathways in macaque monkeys. *Cerebral Cortex*, 17(7), 1561–1569.

Colin, C., Radeau, M., Soquet, A., Demolin, D., Colin, F., & Deltenre, P. (2002). Mismatch negativity evoked by the McGurk–MacDonald effect: A phonetic representation within short-term memory. *Clinical Neurophysiology*, 113(4), 495–506.

Colin, C., Radeau, M., Soquet, A., & Deltenre, P. (2004). Generalization of the generation of an MMN by illusory McGurk percepts: Voiceless consonants. *Clinical Neurophysiology*, 115(9), 1989–2000.

Conrey, B., & Pisoni, D. B. (2006). Auditory-visual speech perception and synchrony detection for speech and nonspeech signals. *Journal of the Acoustical Society of America*, 119(6), 4065.

Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science*, 298(5600), 2013–2015.

Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception*, 9(6), 719–721.

Dodd, B. (1979). Lip reading in infants: Attention to speech presented in- and out-of-synchrony. *Cognitive Psychology*, 11(4), 478–484.

Doesburg, S., Emberson, L., Rahi, A., Cameron, D., & Ward, L. (2008). Asynchrony from synchrony: Long-range gamma-band neural synchrony accompanies perception of audiovisual speech asynchrony. *Experimental Brain Research*, 185(1), 11–20.

Driver, J., & Spence, C. (1998a). Cross–modal links in spatial attention. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 353(1373), 1319–1331.

Driver, J., & Spence, C. (1998b). Crossmodal attention. *Current Opinion in Neurobiology*, 8(2), 245–253.

Driver, J., & Spence, C. (2004). Crossmodal spatial attention: Evidence from human performance. In C. Spence & J. Diver (Eds.), *Crossmodal space and crossmodal attention* (p. 179). Oxford: Oxford University Press.

Dubois, J., Dehaene-Lambertz, G., Perrin, M., Mangin, J.-F., Cointepas, Y., Duchesnay, E., et al. (2008). Asynchrony of the early maturation of white matter bundles in healthy infants: Quantitative landmarks revealed noninvasively by diffusion tensor imaging. *Human Brain Mapping*, 29(1), 14–27.

Erber, N. P. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. *Journal of Speech and Hearing Research*, 12(2), 423–425.

Erber, N. P. (1971). Effects of distance on the visual reception of speech. *Journal of Speech and Hearing Research*, 14(4), 848–857.

Erber, N. P. (1979). Auditory-visual perception of speech with reduced optical clarity. *Journal of Speech & Hearing Research*, 22(2), 212–223.

Falchier, A., Schroeder, C. E., Hackett, T. A., Lakatos, P., Nascimento-Silva, S., Ulbert, I., et al. (2010). Projection from visual areas V2 and prostriata to caudal auditory cortex in the monkey. *Cerebral Cortex*, 20(7), 1529–1538.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1(1), 1–47.

Fendrich, R., & Corballis, P. (2001). The temporal cross-capture of audition and vision. *Attention, Perception, & Psychophysics*, 63(4), 719–725.

Fort, A., & Giard, M.-H. (2004). Multiple electrophysiological mechanisms of audiovisual integration in human perception. In G. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes*. Cambridge, MA: MIT Press.

Foxe, J. J., Morocz, I. A., Murray, M. M., Higgins, B. A., Javitt, D. C., & Schroeder, C. E. (2000). Multisensory auditory–somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Cognitive Brain Research*, 10(1–2), 77–83.

Foxe, J. J., Wylie, G. R., Martinez, A., Schroeder, C. E., Javitt, D. C., Guilfoyle, D., et al. (2002). Auditory-somatosensory multisensory processing in auditory association cortex: An fMRI study. *Journal of Neurophysiology*, 88(1), 540–543.

Froyen, D., Van Atteveldt, N., Bonte, M., & Blomert, L. (2008). Cross-modal enhancement of the MMN to speech-sounds indicates early and automatic integration of letters and speech-sounds. *Neuroscience Letters*, 430(1), 23–28.

Fu, K.-M. G., Johnston, T. A., Shah, A. S., Arnold, L., Smiley, J., Hackett, T. A., et al. (2003). Auditory cortical neurons respond to somatosensory stimulation. *Journal of Neuroscience*, 23(20), 7510–7515.

Gebhard, J. W., & Mowbray, G. H. (1959). On Discriminating the rate of visual flicker and auditory flutter. *The American Journal of Psychology*, 72(4), 521–529.

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *Journal of Cognitive Neuroscience*, 11(5), 473–490.

Givre, S.J., Schroeder, C. E., & Arezzo, J. C. (1994). Contribution of extrastriate area V4 to the surface-recorded flash VEP in the awake macaque. *Vision Research*, 34(4), 415–428.

Grant, K., & Seitz, P. (1998). Measures of auditory–visual integration in nonsense syllables and sentences. *Journal of the Acoustical Society of America*, 104(4), 2438.

Grant, K., & Seitz, P. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *Journal of the Acoustical Society of America*, 108(3), 1197.

Grant, K., & Walden, B. (1996). Evaluating the articulation index for auditory–visual consonant recognition. *Journal of the Acoustical Society of America*, 100(4), 2415.

Grant, K., Walden, B., & Seitz, P. (1998). Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. *Journal of the Acoustical Society of America*, 103(5), 2677.

Green, K. P. (1998). The use of auditory and visual information during phonetic processing: implications for theories of speech perception. In R. Campbell, B. Dodd, D. Burnham (Eds.), *Hearing by eye II: Advances in the psychology of speechreading and auditory-visual speech*. East Sussex, UK: Psychology Press

Guest, S., Catmur, C., Lloyd, D., & Spence, C. (2002). Audiotactile interactions in roughness perception. *Experimental Brain Research*, 146(2), 161–171.

Hackett, T. A., Stepniewska, I., & Kaas, J. H. (1998). Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *The Journal of Comparative Neurology*, 394(4), 475–495.

Hackett, T. A., Smiley, J. F., Ulbert, I., Karmos, G., Lakatos, P., de la Mothe, L. A., & Schroeder, C. E. (2007a). Sources of somatosensory input to the caudal belt areas of auditory cortex. *Perception*, 36(10), 1419–1430.

Hackett, T. A., De La Mothe, L. A., Ulbert, I., Karmos, G., Smiley, J., & Schroeder, C. E. (2007b). Multisensory convergence in auditory cortex, II. Thalamocortical connections of the caudal superior temporal plane. *The Journal of Comparative Neurology*, 502(6), 924–952.

Halle, M., & Stevens, K. N. (1967). Remarks on analysis by synthesis and distinctive features. In W. Wathem-Dunn (ed.), *Models for the perception of speech and visual form*. Cambridge, MA: MIT Press.

Haxby, J., Horwitz, B., Ungerleider, L., Maisog, J., Pietrini, P., & Grady, C. (1994). The functional organization of human extrastriate cortex: A PET-rCBF study of selective attention to faces and locations. *Journal of Neuroscience*, 14(11), 6336–6353.

Hine, T. J., White, A. M., & Chappell, M. (2003). Is there an auditory−visual flash-lag effect? *Clinical & Experimental Ophthalmology*, 31(3), 254–257.

Howard, I. P., & Templeton, W. B. (1966). *Human spatial orientation*. Oxford : John Wiley & Sons.

Jääskeläinen, I. P., Ojanen, V., Ahveninen, J., Auranen, T., Levänen, S., Möttönen, R., et al. (2004). Adaptation of neuromagnetic N1 responses to phonetic stimuli by visual speech in humans. *NeuroReport*, 15(18).

Jancke, L., Mirzazade, S., & Shah, N. J. (1999) Attention modulates activity in the primary and secondary auditory cortex: A functional magnetic resonance imaging study in human subjects. *Neuroscience Letters*, 266, 125–128.

Jones, J. A., & Callan, D. E. (2003). Brain activity during audiovisual speech perception: An fMRI study of the McGurk effect. *NeuroReport*, 14(8), 1129–1133.

Jordan, T. R., & Sergeant, P. (2000). Effects of distance on visual and audiovisual speech recognition. *Language and Speech*, 43(1), 107–124.

Jousmäki, V, & Hari, R. (1998). Parchment-skin illusion: Sound-biased touch. *Current Biology*, 8(6), R190–R191.

Kaiser, J., Hertrich, I., Ackermann, H., Mathiak, K., & Lutzenberger, W. (2005). Hearing lips: Gamma-band activity during audiovisual speech perception. *Cerebral Cortex*, 15(5), 646–653.

Kajikawa, Y., de La Mothe, L., Blumell, S., & Hackett, T. A. (2005). A comparison of neuron response properties in areas A1 and CM of the marmoset monkey auditory cortex: tones and broadband noise. *Journal of Neurophysiology*, 93(1), 22–34.

Kawashima, R., O'Sullivan, B. T., & Roland, P. E. (1995). Positron-emission tomography studies of cross-modality inhibition in selective attentional tasks: Closing the "mind's eye." *Proceedings of the National Academy of Sciences of the USA*, 92(13), 5969–5972.

Kayser, C., Petkov, C. I., Augath, M., & Logothetis, N. K. (2005). Integration of touch and sound in auditory cortex. *Neuron*, 48(2), 373–384.

Kayser, C., Petkov, C. I., Augath, M., & Logothetis, N. K. (2007). Functional imaging reveals visual modulation of specific fields in auditory cortex. *Journal of Neuroscience*, 27(8), 1824–1835.

Kayser, C., Petkov, C. I., & Logothetis, N. K. (2008). Visual modulation of neurons in auditory cortex. *Cerebral Cortex*, 18(7), 1560–1574.

Klucharev, V., Möttönen, R., & Sams, M. (2003). Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. *Cognitive Brain Research*, 18(1), 65–75.

Kosaki, H., Hashikawa, T., He, J., & Jones, E. G. (1997). Tonotopic organization of auditory cortical fields delineated by parvalbumin immunoreactivity in macaque monkeys. *The Journal of Comparative Neurology*, 386(2), 304–316.

Krubitzer, L., Clarey, J., Tweedale, R., Elston, G., & Calford, M. (1995). A redefinition of somatosensory areas in the lateral sulcus of macaque monkeys. *Journal of Neuroscience*, 15(5), 3821–3839.

Kuhl, P., & Meltzoff, A. (1982). The bimodal perception of speech in infancy. *Science*, 218(4577), 1138–1141.

Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., & Schroeder, C. E. (2005). An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *Journal of Neurophysiology*, 94(3), 1904–1911.

Lakatos, P., Chen, C.-M., O'Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron*, 53(2), 279–292.

Leinonen, L., Hyvärinen, J., & Sovijärvi, A. R. A. (1980). Functional properties of neurons in the temporo-parietal association cortex of awake monkey. *Experimental Brain Research*, 39(2), 203–215.

Levänen, S., Jousmäki, V., & Hari, R. (1998). Vibration-induced auditory-cortex activation in a congenitally deaf adult. *Current Biology*, 8(15), 869–872.

Lewald, J., Ehrenstein, W. H., & Guski, R. (2001). Spatio-temporal constraints for auditory–visual integration. *Behavioural Brain Research*, 121(1–2), 69–79.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1), 1–36.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74(6), 431–461.

Liégeois-Chauvel, C., Musolino, A., & Chauvel, P. (1991). Localization of the primary auditory area in man. *Brain*, 114*A*(1), 139–153.

Lipton, M. L., Fu, K.-M. G., Branch, C. A., & Schroeder, C. E. (2006). Ipsilateral hand input to area 3b revealed by converging hemodynamic and electrophysiological analyses in macaque monkeys. *Journal of Neuroscience*, 26(1), 180–185.

Luo, H., Liu, Z., & Poeppel, D. (2010). Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biology*, 8(8), e1000445.

Lütkenhöner, B., Lammertmann, C., Simões, C., & Hari, R. (2002). Magnetoencephalographic correlates of audiotactile interaction. *NeuroImage*, 15(3), 509–522.

Macaluso, E., George, N., Dolan, R., Spence, C., & Driver, J. (2004). Spatial and temporal factors during processing of audiovisual speech: A PET study. *NeuroImage*, 21(2), 725–732.

MacDonald, J, Andersen, S., & Bachmann, T. (2000). Hearing by eye: How much spatial degradation can be tolerated? *Perception*, 29(10), 1155–1168.

MacSweeney, M., Amaro, E., Calvert, G. A., Campbell, R., David, A. S., McGuire, P., et al. (2000). Silent speechreading in the absence of scanner noise: An event-related fMRI study. *NeuroReport*, 11(8).

Martuzzi, R., Murray, M. M., Michel, C. M., Thiran, J.-P., Maeder, P. P., Clarke, S., & Meuli, R. A. (2007). Multisensory interactions within human primary cortices revealed by BOLD dynamics. *Cerebral Cortex*, 17(7), 1672–1679.

Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychology inquiry*. Mahwah, NJ: Lawrence Erlbaum Associates.

Massaro, D. W. (1998). *Perceiving talking faces*. Cambridge, MA: MIT Press.

Massaro, D., Cohen, M., & Smeele, P. (1996). Perception of asynchronous and conflicting visual and auditory speech. *Journal of the Acoustical Society of America*, 100(3), 1777.

Maunsell, J. H., & Gibson, J. R. (1992). Visual response latencies in striate cortex of the macaque monkey. *Journal of Neurophysiology*, 68(4), 1332–1344.

McDonald, J. J., & Ward, L. M. (2000). Involuntary listening aids seeing: Evidence from human electrophysiology. *Psychological Science*, 11(2), 167–171.

McGrath, M., & Summerfield, Q. (1985). Intermodal timing relations and audio–visual speech recognition by normal-hearing adults. *Journal of the Acoustical Society of America*, 77(2), 678–685.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588), 746–748.

Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2000a). Intermodal selective attention in monkeys. I: Distribution and timing of effects across visual areas. *Cerebral Cortex*, 10(4).

Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2000b). Intermodal selective attention in monkeys. II: Physiological mechanisms of modulation. *Cerebral Cortex*, 10(4), 359–370.

Meredith, M., Nemitz, J., & Stein, B. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *Journal of Neuroscience*, 7(10), 3215–3229.

Miller, L. M., & D'Esposito, M. (2005). Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *Journal of Neuroscience*, 25(25), 5884–5893.

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory–visual interactions during early sensory processing in humans: A high-density electrical mapping study. *Cognitive Brain Research*, 14(1), 115–128.

Möttönen, R., Krause, C. M., Tiippana, K., & Sams, M. (2002). Processing of changes in visual speech in the human auditory cortex. *Cognitive Brain Research*, 13(3), 417–425.

Möttönen, R., Schürmann, M., & Sams, M. (2004). Time course of multisensory interactions during audiovisual speech perception in humans: A magnetoencephalographic study. *Neuroscience Letters*, 363(2), 112–115.

Munhall, K., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Attention, Perception, & Psychophysics*, 58(3), 351–362.

Munhall, K., Kroos, C., Jozan, G., & Vatikiotis-Bateson, E. (2004). Spatial frequency requirements for audiovisual speech perception. *Attention, Perception, & Psychophysics*, 66(4), 574–583.

Munhall, K. G., ten Hove, M. W., Brammer, M., & Paré, M. (2009). Audiovisual integration of speech in a bistable illusion. *Current Biology*, 19(9), 735–739.

Murray, M. M., Molholm, S., Michel, C. M., Heslenfeld, D. J., Ritter, W., Javitt, D. C., et al. (2005). Grabbing your ear: Rapid auditory–somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. *Cerebral Cortex*, 15(7), 963–974.

Musacchia G., & Schroeder C. E. (2009) Neuronal mechanisms, response dynamics and perceptual functions of multisensory interactions in auditory cortex. *Hearing Research*, 258, 72–79.

Näätänen, R. (1995). The mismatch negativity: A powerful tool for cognitive neuroscience. *Ear and Hearing*, 16(1).

Nijhawan, R. (1994). Motion extrapolation in catching. *Nature*, 370(6487), 256–257.

Noesselt, T., Rieger, J. W., Schoenfeld, M. A., Kanowski, M., Hinrichs, H., Heinze, H.-J., & Driver, J. (2007). Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. *Journal of Neuroscience*, 27(42), 11431–11441.

Noesselt, T., Bonath, B., Boehler, C. N., Schoenfeld, M. A., & Heinze, H.-J. (2008). On perceived synchrony—neural dynamics of audiovisual illusions and suppressions. *Brain Research*, 1220(0), 132–141.

Ogilvie, J. C. (1956). Effect of auditory flutter on the visual critical flicker frequency. *Canadian Journal of Psychology*, 10(2), 61–68.

Olson, I. R., Gatenby, J. C., & Gore, J. C. (2002). A comparison of bound and unbound audio–visual information processing in the human cerebral cortex. *Cognitive Brain Research*, 14(1), 129–138.

Pallas, S., Roe, A., & Sur, M. (1990). Visual projections induced into the auditory pathway of ferrets. 1. Novel inputs to primary auditory cortex (AI) form the LP/pulvinar complex and the topography of the MGN-AI projection. *Journal of Comparative Neurology*, 298(1), 50–68.

Pandey, P. C., Kunov, H., & Abel, S. M. (1986). Disruptive effects of auditory signal delay on speech perception with lipreading. *Journal of Auditory Research*, 26(1), 27–41.

Passamonti, C., Frissen, I., & Làdavas, E. (2009). Visual recalibration of auditory spatial perception: Two separate neural circuits for perceptual learning. *European Journal of Neuroscience*, 30(6), 1141–1150.

Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science*, 6(2), 191–196.

Paulesu, E., Perani, D., Blasi, V., Silani, G., Borghese, N. A., De Giovanni, U., et al. (2003). A functional-anatomical model for lipreading. *Journal of Neurophysiology*, 90(3), 2005–2013.

Pekkola, J., Ojanen, V., Autti, T., Jääskeläinen, I. P., Möttönen, R., Tarkiainen, A., & Sams, M. (2005). Primary auditory cortex activation by visual speech: An fMRI study at 3 T. *NeuroReport*, 16(2).

Peterson, N. N., Schroeder, C. E., & Arezzo, J. C. (1995). Neural generators of early cortical somatosensory evoked potentials in the awake monkey. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 96(3), 248–260.

Pick, H., Warren, D., & Hay, J. (1969). Sensory conflict in judgments of spatial direction. *Attention, Perception, & Psychophysics*, 6(4), 203–205.

Pilling, M. (2009). Auditory event-related potentials (ERPs) in audiovisual speech perception. *Journal of Speech, Language, and Hearing Research*, 52(4), 1073–1081.

Poeppel, D., Idsardi, W. J., & van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493), 1071–1086.

Radeau, M. (1974). Adaptation au déplacement prismatique sur la base d'une discordance entre la vision et l'audition. *L'Année Psychologique*, 23–33.

Radeau, M., & Bertelson, P. (1987). Auditory-visual interaction and the timing of inputs. *Psychological Research*, 49(1), 17–22.

Raij, T., Uutela, K., & Hari, R. (2000). Audiovisual integration of letters in the human brain. *Neuron*, 28(2), 617–625.

Rauschecker, J. P., Tian, B., Pons, T., & Mishkin, M. (1997). Serial and parallel processing in rhesus monkey auditory cortex. *The Journal of Comparative Neurology*, 382(1), 89–103.

Reale, R. A., Calvert, G. A., Thesen, T., Jenison, R. L., Kawasaki, H., Oya, H., et al. (2007). Auditory-visual processing represented in the human superior temporal gyrus. *Neuroscience*, 145(1), 162–184.

Recanzone, G. H. (1998). Rapidly induced auditory plasticity: The ventriloquism aftereffect. *Proceedings of the National Academy of Sciences of the USA*, 95(3), 869–875.

Recanzone, G. H. (2003). Auditory influences on visual temporal rate perception. *Journal of Neurophysiology*, 89(2), 1078–1093.

Recanzone, G. H., Guard, D. C., & Phan, M. L. (2000). Frequency and intensity response properties of single neurons in the auditory cortex of the behaving macaque monkey. *Journal of Neurophysiology*, 83(4), 2315–2331.

Reisberg, D., McLean, J., & Goldfield, A. (1987). Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 97–113). Hillsdale, NJ: Lawrence Erlbaum Associates.

Robinson, C. J., & Burton, H. (1980). Organization of somatosensory receptive fields in cortical areas 7b, retroinsula, postauditory and granular insula of M. fascicularis. *Journal of Comparative Neurology*, 192(1), 69–92.

Rockland, K. S., & Pandya, D. N. (1979). Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. *Brain Research*, 179(1), 3–20.

Roe, A., Pallas, S., Hahm, J., & Sur, M. (1990). A map of visual space induced in primary auditory cortex. *Science*, 250(4982), 818–820.

Rosenblum, L., & Saldãna, H. M. (1996). An audiovisual test of kinematic primitives for visual speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 22(2), 318–331.

Rosenblum, L., Schmuckler, M., & Johnson, J. (1997). The McGurk effect in infants. *Attention, Perception, & Psychophysics*, 59(3), 347–357.

Rosenthal, O., Shimojo, S., & Shams, L. (2009). Sound-induced flash illusion is resistant to feedback training. *Brain Topography*, 21(3), 185–192.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex*, 17(5), 1147–1153.

Sams, M., Aulanko, R., Hämäläinen, M., Hari, R., Lounasmaa, O. V., Lu, S.-T., & Simola, J. (1991). Seeing speech: Visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters*, 127(1), 141–145.

Sato, M., Baciu, M., Lœvenbruck, H., Schwartz, J.-L., Cathiard, M.-A., Segebarth, C., & Abry, C. (2004). Multistable representation of speech forms: A functional MRI study of verbal transformations. *NeuroImage*, 23(3), 1143–1151.

Schmolesky, M. T., Wang, Y., Hanes, D. P., Thompson, K. G., Leutgeb, S., Schall, J. D., & Leventhal, A. G. (1998). Signal timing across the macaque visual system. *Journal of Neurophysiology*, 79(6), 3272–3278.

Schorr, E. A., Fox, N. A., van Wassenhove, V., & Knudsen, E. I. (2005). Auditory-visual fusion in speech perception in children with cochlear implants. *Proceedings of the National Academy of Sciences of the USA*, 102(51), 18748–18750.

Schroeder, C. E., & Foxe, J. J. (2002). The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Cognitive Brain Research*, 14(1), 187–198.

Schroeder, C. E., Seto, S., Arezzo, J. C., & Garraghty, P. E. (1995). Electrophysiological evidence for overlapping dominant and latent inputs to somatosensory cortex in squirrel monkeys. *Journal of Neurophysiology*, 74(2), 722–732.

Schroeder, C. E., Seto, S., & Garraghty, P. E. (1997). Emergence of radial nerve dominance in median nerve cortex after median nerve transection in an adult squirrel monkey. *Journal of Neurophysiology*, 77(1), 522–526.

Schroeder, C E, Mehta, A. D., & Givre, S. J. (1998). A spatiotemporal profile of visual system activation revealed by current source density analysis in the awake macaque. *Cerebral Cortex*, 8(7), 575–592.

Schroeder, C.E., Lindsley, R. W., Specht, C., Marcovici, A., Smiley, J. F., & Javitt, D. C. (2001). Somatosensory input to auditory association cortex in the macaque monkey. *Journal of Neurophysiology*, 85(3), 1322–1327.

Schroeder, C. E., Molhom, S., Lakatos, P., Ritter, W., & Foxe, J. J. (2004). Human–simian correspondence in the early cortical processing of multisensory cues. *Cognitive Processing*, 5(3), 140–151.

Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends in Cognitive Sciences*, 12(3), 106–113.

Schürmann, M., Caetano, G., Hlushchuk, Y., Jousmäki, V., & Hari, R. (2006). Touch activates human auditory cortex. *NeuroImage*, 30(4), 1325–1331.

Schwartz, J. L., Robert-Ribes, J., & Escudier, P. (1998). Ten years after Summerfield: A taxonomy of models for audio-visual fusion in speech perception. In R. Campbell, B. Dodd, & D. Burnham (Eds.), *Hearing by eye II: Advances in the psychology of speechreading and auditory-visual speech* (pp. 85–108). East Sussex, UK: Psychology Press.

Sekiyama, K., Kanno, I., Miura, S., & Sugita, Y. (2003). Auditory-visual speech perception examined by fMRI and PET. *Neuroscience Research*, 47(3), 277–287.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions: What you see is what you hear. *Nature*, 408(6814), 788.

Shams, L., Iwaki, S., Chawla, A., & Bhattacharya, J. (2005). Early modulation of visual cortex by sound: an MEG study. *Neuroscience Letters*, 378(2), 76–81.

Shipley, T. (1964). Auditory flutter-driving of visual flicker. *Science*, 145(3638), 1328–1330.

Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., & Small, S. L. (2007). Hearing Lips and Seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex*, 17(10), 2387–2399.

Smiley, J. F., Hackett, T. A., Ulbert, I., Karmas, G., Lakatos, P., Javitt, D. C., & Schroeder, C. E. (2007). Multisensory convergence in auditory cortex, I. Cortical connections of the caudal superior temporal plane in macaque monkeys. *The Journal of Comparative Neurology*, 502(6), 894–923.

Soto-Faraco, S., Lyons, J., Gazzaniga, M., Spence, C., & Kingstone, A. (2002). The ventriloquist in motion: Illusory capture of dynamic information across sensory modalities. *Cognitive Brain Research*, 14(1), 139–146.

Soto-Faraco, S., Navarra, J., & Alsius, A. (2004). Assessing automaticity in audiovisual speech integration: Evidence from the speeded classification task. *Cognition*, 92(3), B13–B23.

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge, MA: MIT Press.

Stein, B. E., London, N., Wilkinson, L. K., & Price, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis. *Journal of Cognitive Neuroscience*, 8(6), 497–506.

Steinschneider, M., Tenke, C. E., Schroeder, C. E., Javitt, D. C., Simpson, G. V., Arezzo, J. C., & Vaughan Jr., H. G. (1992). Cellular generators of the cortical auditory evoked potential initial component. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 84(2), 196–200.

Steinschneider, M., Schroeder, C. E., Arezzo, J. C., & Vaughan, H. G., Jr. (1994). Speech-evoked activity in primary auditory cortex: Effects of voice onset time. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 92(1), 30–43.

Steinschneider, M., Reser, D., Fishman, Y., Schroeder, C., & Arezzo, J. (1998). Click train encoding in primary auditory cortex of the awake monkey: Evidence for two mechanisms subserving pitch perception. *Journal of the Acoustical Society of America*, 104(5), 2935.

Stekelenburg, J. J., & Vroomen, J. (2005). An event-related potential investigation of the time-course of temporal ventriloquism. *NeuroReport*, 16(6).

Stekelenburg, J. J., Vroomen, J., & de Gelder, B. (2004). Illusory sound shifts induced by the ventriloquist illusion evoke the mismatch negativity. *Neuroscience Letters*, 357(3), 163–166.

Stratton, G. M. (1897). Vision without inversion of the retinal image. *Psychological Review*, 4(4), 341–360.

Summerfield, Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 3–51). Hillsdale, NJ: Lawrence Erlbaum Associates.

Summerfield, Q. (1992). Lipreading and audio-visual speech perception. *Philosophical Transactions: Biological Sciences*, 335(1273), 71–78.

Summerfield, Q., & McGrath, M. (1984). Detection and resolution of audio-visual incompatibility in the perception of vowels. *The Quarterly Journal of Experimental Psychology Section A*, 36(1), 51–74.

Sur, M, Garraghty, P., & Roe, A. (1988). Experimentally induced visual projections into auditory thalamus and cortex. *Science*, 242(4884), 1437–1441.

Sur, M., Pallas, S. L., & Roe, A. W. (1990). Cross-modal plasticity in cortical development: differentiation and specification of sensory neocortex. *Trends in Neurosciences*, 13(6), 227–233.

Thomas, G. J. (1941). Experimental study of the influence of vision on sound localization. *Journal of Experimental Psychology*, 28, 167–177.

Tiippana, K., Andersen, T. S., & Sams, M. (2004). Visual attention modulates audiovisual speech perception. *European Journal of Cognitive Psychology*, 16(3), 457–472.

van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron*, 43(2), 271–282.

van Atteveldt, N., Roebroeck, A., & Goebel, R. (2009). Interaction of speech and script in human auditory cortex: Insights from neuro-imaging and effective connectivity. *Hearing Research*, 258(1–2), 152–164.

van Atteveldt, N., Blau, V., Blomert, L., & Goebel, R. (2010). fMRI-adaptation indicates selectivity to audiovisual content congruency in distributed clusters in human superior temporal cortex. *BMC Neuroscience*, 11(1), 11.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the USA*, 102(4), 1181–1186.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, 45(3), 598–607.

Vidal, J., Giard, M.-H., Roux, S., Barthélémy, C., & Bruneau, N. (2008). Cross-modal processing of auditory–visual stimuli in a no-task paradigm: A topographic event-related potential study. *Clinical Neurophysiology*, 119(4), 763–771.

von Schiller, P. (1932). Die rauhigkeit als intermodale erscheinung. *Zeitschrift für Psychologie Bildung*, 127, 265–289.

Vroomen, J., & de Gelder, B. (2004). Temporal ventriloquism: sound modulates the flash-lag effect. *Journal of Experimental Psychology: Human Perception and Performance*, 30(3), 513–518.

Vroomen, J., Bertelson, P., & De Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention. *Attention*, *Perception*, *& Psychophysics*, 63(4), 651–659.

Watkins, S., Shams, L., Tanaka, S., Haynes, J.-D., & Rees, G. (2006). Sound alters activity in human V1 in association with illusory visual perception. *NeuroImage*, 31(3), 1247–1256.

Woods, T. M., & Recanzone, G. H. (2004). Visually induced plasticity of auditory spatial perception in macaques. *Current Biology*, 14(17), 1559–1564.

Wright, T. M., Pelphrey, K. A., Allison, T., McKeown, M. J., & McCarthy, G. (2003). Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cerebral Cortex*, 13(10), 1034–1043.

Yau, J. M., Olenczak, J. B., Dammann, J. F., & Bensmaia, S. J. (2009). Temporal frequency channels are linked across audition and touch. *Current Biology*, 19(7), 561–566.

Chapter 12
Redefining the Functional Organization of the Planum Temporale Region: Space, Objects, and Sensory–Motor Integration

Gregory Hickok and Kourosh Saberi

12.1 Introduction: Definitions, History, and Preview

The planum temporale (PT) is defined anatomically by the triangular surface on the supratemporal plane posterior to Heschl's gyrus (see Clarke and Morosan, Chapter 2). Its posterior boundary is the termination point of the Sylvian fissure, its medial boundary is the insula or the point where the supratemporal plane transitions into the parietal operculum, and its lateral boundary is the lateral lip of the superior temporal gyrus (Fig. 12.1).

Interest in the PT was promoted by the discovery of a left–right asymmetry in this structure, with the left PT being larger than the right in most (~65%) right-handed individuals (Geschwind & Levitsky, 1968). Given that the left PT comprises a part of classical Wernicke's area, the relatively larger size was thought to be an anatomical reflection of the region's functional specialization for speech processing.

Several subsequent findings cast doubt on this view. One was that PT asymmetry was found to correlate with a nonspeech function, musical ability (Schlaug et al., 1995). Another was that a similar leftward PT asymmetry was found in chimpanzees, a species without speech ability at all (Gannon et al., 1998; this of course questions the basis of the association with musical ability as well!). A third was that structural asymmetries of the PT did not correlate with language dominance as assessed directly using the intracarotid sodium amytal (Wada) procedure (Dorsaint-Pierre et al., 2006).

Results from functional imaging corroborated these findings. It was reported, for example, that the left PT responded equally well or even more robustly during processing of tone stimuli compared to speech (Binder et al., 1996; Hickok et al., 2003). In fact, a range of nonspeech signals were found to activate the PT including

G. Hickok (✉) • K. Saberi
Department of Cognitive Sciences, University of California, Irvine, CA 92697, USA
e-mail: Greg.Hickok@uci.edu; Saberi@uci.edu

D. Poeppel et al. (eds.), *The Human Auditory Cortex*, Springer Handbook of Auditory Research 43, DOI 10.1007/978-1-4614-2314-0_12,

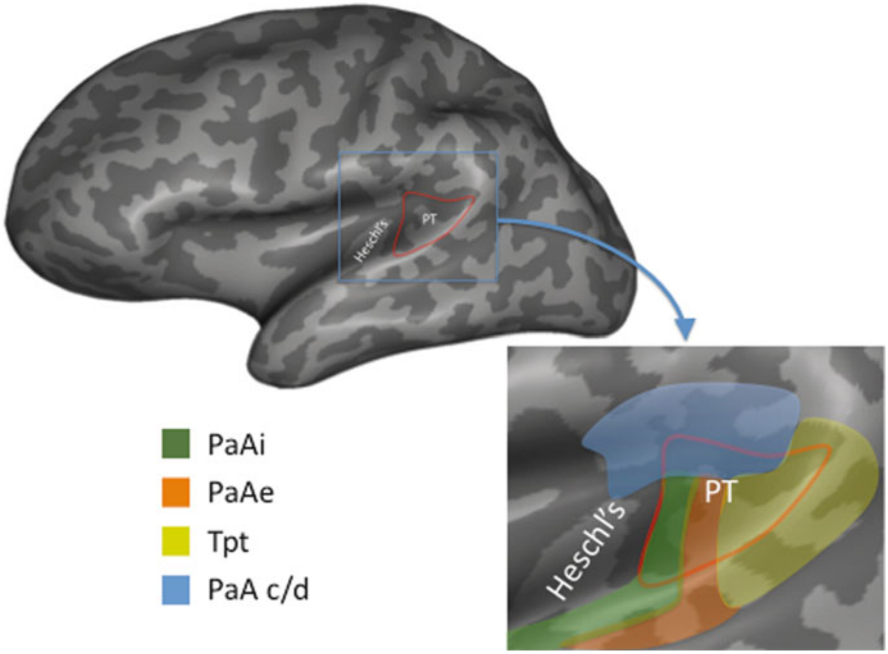


Fig. 12.1 Location and cytoarchitectonic organization of the planum temporale. The location of the planum temporale on the posterior supratemporal plane is indicated in red outline on an inflated representation of the brain that shows structures buried in sulci and fissures. The inset shows a close up of the planum temporale region. Colors indicate approximate location of different cytoarchitectonic fields as delineated by Galaburda and Sanides (1980). Note that there are four different fields within the planum temporale, suggesting functional differentiation, and that these fields extend beyond the planum temporale. The area in yellow corresponds to cytoarchitectonic area Tpt which is not considered part of auditory cortex proper. Functional area Spt likely falls within cytoarchitectonic area Tpt, although this has never been directly demonstrated. PaAi, parakoniocortex—internal; PaAe, parakoniocortex—external; PaA c/d, parakoniocortex—caudodorsal Tpt, Temporal–parietal

multiple spatially dispersed sound sources (Zatorre et al., 2002; Smith et al., 2009), moving or spatially changing sound sources (Warren et al., 2002; Smith et al., 2004), visual speech (Calvert et al., 1997; Okada & Hickok, 2009), and auditory–motor integration (Buchsbaum et al., 2001; Wise et al., 2001; Hickok et al., 2003; Overath et al., 2007; Hickok et al., 2009).

This heterogeneity of function has led some authors to the view that the PT supports a general computation that operates over many classes of stimulus types. On one variant, the PT functions as a "computational hub" that takes as input a range of acoustic signals, performs a pattern matching operation, and then channels its output according to the nature of the signal; speech and other auditory objects would be channeled into one pathway, spatial information into another, and so on (Griffiths & Warren, 2002). According to another variant, which emphasizes auditory dorsal stream function (Rauschecker & Scott, 2009), the posterior superior temporal region (which includes the PT) supports the implementation of "internal models," mechanisms that model the input/output characteristics of the motor system for the purpose of motor control and/or sensory prediction (forward models). They argue for a "common computation mechanism" that implements internal models not just for speech as has been proposed previously (Guenther et al., 1998; van Wassenhove et al., 2005), but also for spatial hearing-related functions.

There is an alternative conceptualization of the PT, however. Although the computational hub hypothesis interprets the PT's functional heterogeneity as evidence for some common computation that operates over a range of stimulus types and that characterizes the function of the whole structure, another possibility is that the PT's heterogeneity of function is evidence for heterogeneity of function. That is, perhaps the PT is not functionally homogeneous but instead is composed of subfields that perform different operations, for example, spatial versus sensory–motor processes.

A terminological note: Although it is common to refer to the PT as if it were a functional region, this is misleading as functional boundaries do not respect gross anatomical boundaries. When referring specifically to the planum temporale as it is anatomically defined (e.g., if we are discussing previous studies that mapped the response patterns within this region), the term PT is used in this chapter. However, when referring to the broader functional–anatomic region, which likely spans portions of the parietal operculum, lateral portions of the superior temporal gyrus, and even the superior temporal sulcus, the term, *planum temporale region (PTR)* is used here.

12.2 Cytoarchitectonics of the Planum Temporale Region

Cytoarchitectonic studies of human auditory cortex demonstrate that several subregions exist within the PT and that the boundaries of these cytoarchitectonic fields extend beyond the gross anatomical boundaries of the PT (Fig. 12.1, see also Clarke and Morosan, Chapter 2). Galaburda and Sanides (Galaburda & Sanides, 1980) identified four areas that are at least partly within the PT. Three are classified as parakoniocortex, cortical fields with prominent granularity in layer IV and relatively sparse layer V, but to a lesser degree than the extremely granular konio fields, which are found on Heschl's gyrus. PaAi (parakoniocortex—internal) is just lateral/posterior to Heschl's gyrus and corresponds to the lateral belt region; PaAe (parakoniocortex—external) is lateral/posterior to PaAi and corresponds to the parabelt region; PaA c/d (parakoniocortex —caudodorsal) is caudodorsal to Hechl's gyrus. The fourth area, Tpt (temporal–parietal), which occupies much of the posterior portion of the PT, has a weak layer IV and more prominent layer V and so is not classified as parakoniocortex. Galaburda and Sanides emphasize that Tpt "lacks specialty features of sensory cortex" (p. 609) and so should not be considered part of auditory cortex. This conclusion is reinforced by comparative studies that indicate that the homologous area in monkey, also called Tpt, is not considered part of auditory cortex (Sweet et al., 2005; Smiley et al., 2007). Interestingly, much of the anatomical asymmetry in the PT appears to be found in Tpt (Galaburda & Sanides, 1980). As is clear from Fig. 12.1, each of these regions extend well beyond the boundaries of the PT to include portions of the superior temporal sulcus, parietal operculum, and even supramarginal gyrus. A similar organization of the human PT was reported by Sweet et al. using other histological stains (Sweet et al., 2005).

These findings indicate that the PT (and more broadly, the PTR) is anatomically heterogeneous, including belt and parabelt auditory fields, as well as an area, Tpt, which cannot be characterized as auditory cortex.

12.3 Role of the PT in Auditory Space and Object Processing

Portions of the PT have been found to respond to spatial auditory signals including moving sound sources (Baumgart et al., 1999; Warren et al., 2002; Smith et al., 2004) and nonmoving but spatially varying sounds (Warren & Griffiths, 2003; Smith et al., 2004; Smith et al., 2007). Several studies have assessed the relative selectivity of these spatial responses. One line of studies investigated whether responses in the PT were motion selective by contrasting moving sounds with sound sources that are perceived to jump from one location to the next (spatial change but without perceived motion). Two studies found the same degree of activation in the PT for both conditions, arguing against the view that the PT contains a motion-dedicated cortical region, analogous to visual area MT (Smith et al., 2004, 2007).

Another line of studies assessed the relative spatial selectivity of PT responses compared to nonspatial signals such as pitch or environmental sounds. This work has shown convincingly that sequences of spatially varying sound sources (changes in the location) yield greater activity in the PT than sequences of non-spatial variation (e.g., changes in pitch); the latter produce greater activation in Heschl's gyrus and more anterior auditory fields (Warren & Griffiths, 2003; Barrett & Hall, 2006; Altmann et al., 2007). Observations such as these have been used to argue for the existence of anterior "what" and posterior "where" pathways in human auditory cortex with the PT a major structure in the "where" pathway (Warren & Griffiths, 2003; Altmann et al., 2007; Rauschecker & Scott, 2009).

But other studies have cast doubt on the idea of a pure "where" function within the PT, or anywhere in cortex. Use PET, Zatorre et al. (2002), for example, found that increasing the number of sound source locations correlated with PT activity only when spatial information was useful for auditory object segregation. Specifically, presenting a noise stimulus at 1, 2, 3, 4, or 6 locations (in sequence) did not correlate with changes in PT activity, but presenting a set of 12 environmental sounds *simultaneously* at either 1 location or distributed over 2, 3, 4, or 6 locations did correlate positively with PT activity (Zatorre et al., 2002). As noted in the preceding text, however, other studies have reported a modulation of PT activity with spatial manipulations alone, which appears to contradict the result of Zatorre et al.. A more recent study using functional magnetic resonance imaging (fMRI; Smith et al., 2009) clarified the situation. This study manipulated the number of auditory objects, in this case with speech stimuli (1 vs. 3 talkers), and the number of spatial locations at which the stimuli

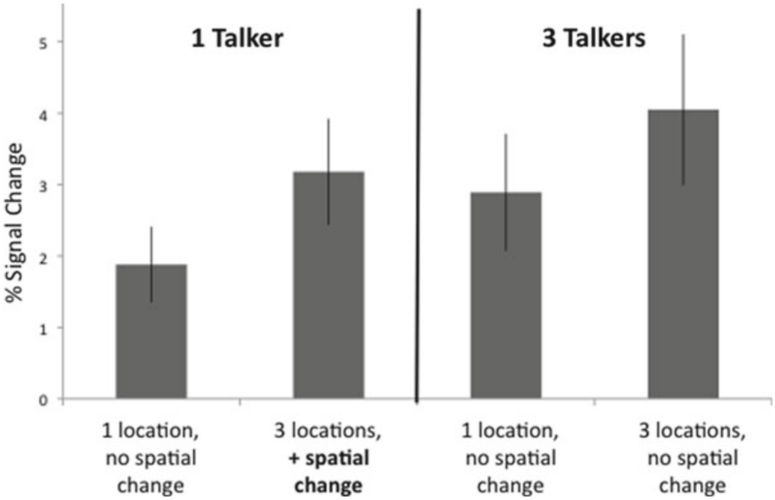


Fig. 12.2 FMRI signal change in a region of interest (ROI) in the human planum temporale defined by a spatial manipulation. The ROI was defined by contrasting a spatial change (a single talker's voice that bounced between three spatial locations) with a no-spatial change condition (a single talker's voice that was stationary at one spatial location). The spatial effect is evident in the left two bars with spatial change eliciting more activity than no spatial change; this is the standard effect observed in the PT. Note, however, equal or greater signal amplitude is observed with no spatial change, that is, by simply adding talkers either in one location or in three different (static) locations. (Adapted from Smith et al., 2009.)

were presented (1 vs. 3 locations). Consistent with previous reports of "pure" spatial effects, presenting a single speech stimulus (talker) from one location yielded less activation than a single speech stimulus that bounced between three locations (Fig. 12.2, left half of graph). Spatial variation clearly modulated the response of the PT. However, simply adding talkers to the speech stream (three talkers presented simultaneously at one location) resulted in a similar increase in activity level in the PT (cf. middle two bars in Fig. 12.2). Finally, presenting three talkers simultaneously at three locations, *without spatial change*, resulted in the highest activity level (Fig. 12.2, right most bar). Thus, spatial change alone can modulate PT activity, but, consistent with the finding of Zatorre et al. the response in this region is particularly sensitive to the interaction of spatial and auditory object manipulations. As Zatorre et al. suggested, this finding is consistent with the proposal that portions of the PT are involved in auditory stream segregation and may use spatial cues in this service (Zatorre et al., 2002; Smith et al., 2009). On this view, there is no dedicated "where" stream in auditory cortex and "spatial" responses reflect a spatial *contribution* to some other computation (Middlebrooks, 2002). See also later.

In sum, although there is ample evidence that portions of the PT respond to spatial manipulations, there is no evidence to date suggesting that the PT contains a region dedicated to computing spatial location and/or auditory motion. Instead, spatial responses may reflect the *use* of spatial information for other functions.

12.4 Role of the PTR in Auditory–Motor Integration

Auditory–motor integration is critical for several aspects of speech and auditory processing. In the speech domain, it is well documented that auditory feedback has relatively rapid (~100 ms) effects on speech production, for example, the disruptive effects of delayed auditory feedback (Yates, 1963; Stuart et al., 2002) and the pitch- or F1-shift reflex (Burnett et al., 1998; Houde & Jordan, 1998; Tourville et al., 2008). Auditory–motor integration is also critical in development where the young child must use auditory information in his or her linguistic environment to guide articulatory processes that are aimed at reproducing those sounds with the vocal tract (Hickok & Poeppel, 2000, 2004, 2007). This requirement extends to the suprasegmental domain: not only does one have to learn how to produce the individual sounds of the language, but also the sequences of sounds and syllables that correspond to the words of the language (Hickok & Poeppel, 2007). The ability to reproduce nonlinguistic sounds and sequences with the vocal tract (e.g., a dog's bark or a melody) demonstrates that auditory–motor integration is not restricted to the speech domain.

Several lines of evidence link auditory–motor integration with the left posterior PTR. Damage to this area is associated with conduction aphasia (Fig. 12.3A) (Buchsbaum et al., e-pub 2011), a syndrome that results in a speech *production* deficit in which a patient's speech output is fluent but marked by abundant phonemic errors in spontaneous speech (Benson et al., 1973; Goodglass, 1992). Conduction aphasics also have difficulty with verbatim repetition of speech, which is exacerbated when speech has little semantic content (Goodglass, 1992). Receptive speech abilities are largely preserved however, even for speech they cannot repeat (Baldo et al., 2008). The preserved receptive speech and fluent speech output suggest that the deficit in conduction aphasia involves neither acoustic perception nor motor execution of speech, but rather the interface of these two systems (Hickok et al., 2011; Buchsbaum et al., e-pub 2011). Direct cortical stimulation of the PTR has been reported to induce symptoms of conduction aphasia (Anderson et al., 1999).

Functional imaging studies of auditory–motor tasks have similarly implicated the left posterior PTR (Buchsbaum et al., 2001; Wise et al., 2001; Hickok et al., 2009). A series of studies have identified a set of cortical areas that have auditory–motor response properties, responding both during the perception and production of speech in verbatim repetition tasks (covert, i.e., subvocal speech is used in these studies to eliminate the auditory response to hearing one's own voice during repetition) (Paus et al., 1996; Hickok et al., 2003; Hickok et al., 2009). The auditory–motor network identified by these studies includes posterior frontal regions (pars opercularis/area 44 of Broca's area as well as more dorsal premotor regions), the superior temporal sulcus bilaterally, and the left posterior PTR (Fig. 12.3B). This posterior PTR activation likely falls within the distribution of cytoarchitectonic area Tpt but appears to be quite focal in most individuals and therefore probably comprises a subset of Tpt. This functionally defined area in the posterior PTR has been termed Spt (Sylvian–parietal–temporal) to distinguish it from the anatomically

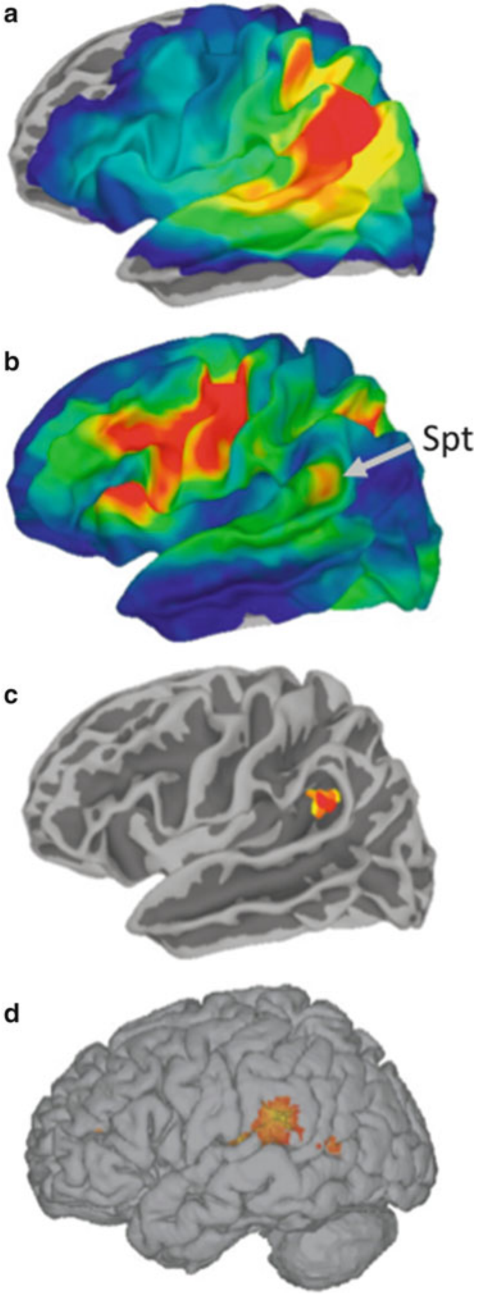


Fig. 12.3 (**a**) Distribution of lesions associated with conduction aphasia (*n* = 16). Warmer colors indicate greater overlap. (**b**) Location of area Spt as identified in a listen and rehearse fMRI paradigm (*n* = 106). (**c**) Overlap between maximal density of lesions associated with conduction aphasia and fMRI localization of Spt. (**a–c** from Buchsbaum et al., 2011). (**d**) fMRI localization of the effect of altered auditory feedback minus unaltered feedback. (Adapted from Tourville et al., 2008.)

defined area Tpt (Hickok et al., 2003). As Spt is strongly left dominant, it is worth noting again that of the cytoarchitectonic areas in the PT, Tpt exhibits the greatest degree of leftward asymmetry (Galaburda et al., 1978; Galaburda & Sanides, 1980), further reinforcing the link between Spt and Tpt. The location of Spt appears to overlap substantially with the region most consistently damaged in conduction aphasia (Fig. 12.3C) (Buchsbaum et al., e-pub 2011).

Beyond auditory–motor response properties, Spt exhibits several features characteristic of sensory–motor integration areas that have been identified in monkey parietal cortex (Andersen, 1997; Colby & Goldberg, 1999). For example, Spt appears to have both sensory-weighted and motor-weighted cell populations as evidenced by multivariate pattern analysis of fMRI data that has found distinguishable patterns of activity within Spt during the sensory and motor phases of a sensory–motor task (Hickok et al., 2009). Spt is not speech specific, responding equally well during the perception and covert production (humming) of melodic tone sequences (Hickok et al., 2003). However, like sensory–motor areas in the monkey parietal lobe, Spt does show motor effector specificity, responding more when the motor task involves the vocal tract (humming) than when it involves the manual articulators (imagined piano playing) despite identical sensory input (Pa & Hickok, 2008). This collection of observations has led to the proposal that Spt, rather than being an *auditory*–motor interface area, is a *sensory*–motor interface area for the vocal tract effector system (Pa & Hickok, 2008; Hickok et al., 2009). This is consistent with Spt's presumed location within nonauditory area Tpt. It is also relevant in this context that area Tpt is substantially more developed in humans than in monkeys (Galaburda et al., 1978). This may reflect the dramatically increased load on sensory–motor coordination of vocal tract actions with the evolution of speech.

The aforementioned studies utilize sequences of sounds to study auditory–motor interaction. At least one study (Tourville et al., 2008) used an altered auditory feedback paradigm in which subjects phonated a vowel under conditions of normal or altered feedback (F1 shift). An activation focus was found in the PTR that responded more during altered than unaltered feedback (Fig. 12.3D) suggesting that this region supports sensory–auditory–motor interaction on multiple scales, that is, both at the level of phonetic features and sound sequences. It is an open question whether these levels rely on the same computational network or on parallel circuits.

Spt has been characterized as an auditory–motor integration area, but what does this mean computationally? There are two hypotheses. One is that the region that includes Spt, as well as the STG more broadly, comprises an auditory target map that compares the predicted auditory consequences of a speech act (a forward prediction) with the actual auditory feedback and generates an error signal in cases of mismatch (Golfinopoulos et al., 2010). Evidence for this view comes from the observation that the STG, including PTR, is more strongly activated when the subject's speech output is altered compared to when it is not (Christoffels et al., 2007; Tourville et al., 2008; Takaso et al., 2010). Another possibility, hypothesized to hold of Spt specifically, is that it is performing a coordinate transform between auditory-based representations and a motor-based representations of speech (Hickok et al., 2009, 2011). Evidence for this claim comes from neuropsychology: the pattern of sparing and loss in conduction aphasia has been characterized as a disconnection between intact auditory and motor speech systems (Jacquemot et al., 2007; Buchsbaum et al., e-pub 2011) and the lesions in this syndrome implicate Spt (Buchsbaum et al., e-pub 2011)

12.5 Functional Subdivision of the PTR into Auditory versus Sensory–Motor Function

It has recently been proposed (Hickok, 2009) that the PT is subdivided—or more accurately, the PTR—into at least two broad regions, an anterior sector that corresponds to unimodal auditory cortex and a posterior sector, area Tpt, that is more multimodal (Hackett et al., 2007; Zheng et al., 2009,) including a region, Spt, that specifically supports sensory–motor integration for vocal tract actions (Hickok et al., 2009). The cytoarchitectonic data reviewed in the preceding text supports this view in that the anterior portion of the PT has been classified as unimodal auditory cortex, whereas the posterior sector, area Tpt, lacks the defining features of sensory cortex. It does share some similarities, however, with area 44 (the pars opercularis) in Broca's region. As Galaburda puts it, Tpt."..exhibits a degree of specialization like that of Area 44 in Broca's region. … Thus 44 and Tpt are equivalent transitional areas between the paramotor and the generalized cortices of the prefrontal area, and between parakoniocortex and temporoparietal occipital junction areas respectively. …the intimate relationship and similar evolutionary status of Areas 44 and Tpt allows for a certain functional overlap." (Galaburda, 1982, pp. 442–443). As noted previously, area 44 is part of the sensory–motor integration circuit that includes Spt in the posterior PTR. These findings are consistent with the view that the posterior PTR supports sensory–motor functions and is distinct from more anterior fields in the PTR.

Given this anatomical distinction, one wonders whether the spatial-related functions associated with the PT involve more anterior regions than the sensory–motor functions. A recent within-subject fMRI study addressed this question directly (Isenberg et al., 2011). This study employed a sensory–motor task (speech shadowing: immediately repeating back heard speech) as well as an auditory motion condition. In both individual-subject analyses and in the averaged group data, the activations for the sensory–motor and auditory motion conditions were distinct and in posterior versus anterior regions of the PT, respectively (Fig. 12.4).

12.5.1 The PTR in the Context of the Dorsal and Ventral Auditory Streams

There is convergence on the view that the PT is part of the auditory dorsal stream (Warren et al., 2005; Hickok & Poeppel, 2007; Rauschecker & Scott, 2009) but less consensus regarding its function. The dominant competing theories are the

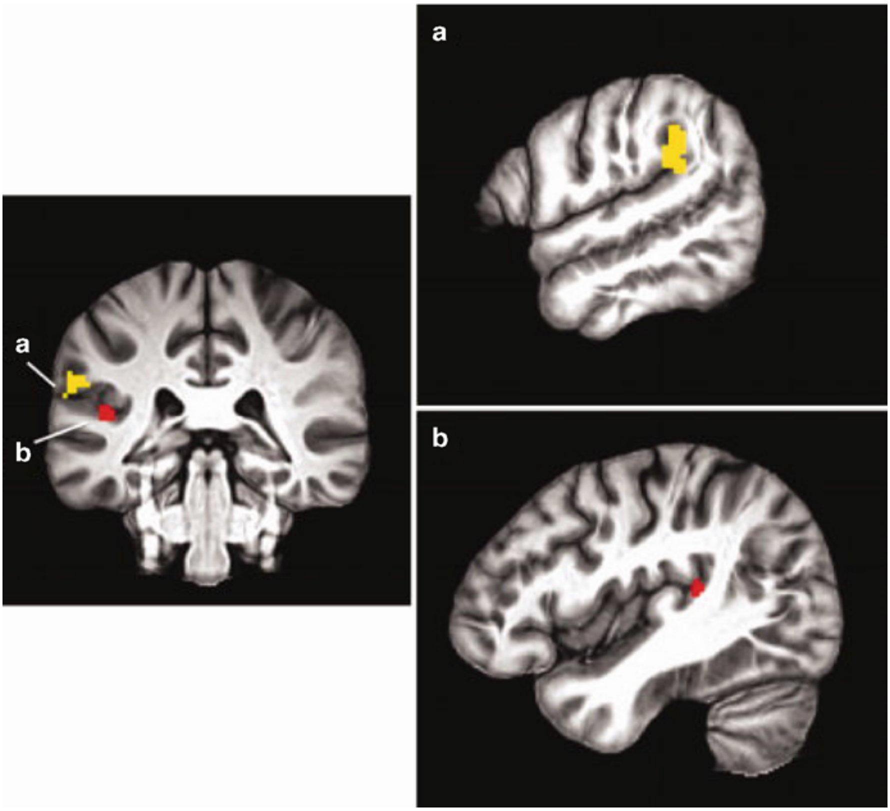


Fig. 12.4 fMRI activation for sensory-motor task is shown in yellow **(a)** and the spatial hearing manipulation is shown in red **(b)**. Adapted from from Isenberg et al., 2011.)

sensory–motor theory (Hickok & Poeppel, 2007) and the spatial "where" theory (Rauschecker, 1998; Rauschecker & Scott, 2009). In previous sections we summarized the evidence for the sensory–motor theory as well as the evidence for and against a pure "where" theory. Here we describe a reinterpretation of the notion of sensory processing streams motivated in part by the functionally subdivided model of the PT.

Whereas most research in the auditory system emphasizes stimulus characteristics, for example, spatial versus pitch variation, we suggest that an emphasis on the behavioral goal (task) may be more important in some instances and can clarify issues in the debate over the function of different processing streams in auditory cortex. In theorizing about dorsal and ventral streams within the visual system there has been a shift of focus from stimulus-centered ideas (form vs. space; Ungerleider & Mishkin, 1982) to goal or task-centered ideas (recognition vs. sensory–motor interaction; Milner & Goodale, 1995). This same shift of focus has been emphasized by some authors in the auditory domain (Hickok & Poeppel, 2000, 2007; Warren et al., 2005).

This shift of focus onto the behavioral goals of a task can be generalized to provide a framework for thinking about sensory processing streams. Consider an example from the spatial domain. A spatially localizable signal has certain sensory features—interaural time difference, interaural level difference, and the particular filtering properties of the outer ear—that can be used to compute location information. However, spatial information can be put to use in a variety of ways. For example, in addition to informing explicit localization decision tasks, spatial information can drive auditory stream segregation (Bregman, 1990) or any number of sensory–motor processes, such as orienting, tracking, approach, or avoidance responses. Note that the goals (effects or output) of these tasks are very different. In auditory stream segregation the goal is to resolve an auditory object, such as a single voice in a noisy room. This is arguably a "ventral stream" function in that the goal is to identify *what* an object is (what is this person saying?). In a sensory–motor process the goal may be to generate a motor command such as a saccade or a head movement or locomotion toward or away from the sound source. And in explicit localization judgments the goal may be to make a spatial decision such as whether one sound occurred in the same or different location than a previous sound. Assuming that different goals (stream segregation, saccade generation, location/motion decision) implicate different neural systems, spatial auditory information must enter into a range of task-dependent, distinct processes. A similar argument could be made for a feature such as frequency (pitch) which could be used for stream segregation, sensory–motor integration (mimicking a tone via humming or reproducing it on a musical instrument), voice identification, explicit pitch discrimination decisions, and so on.

So the same sensory cues can be used for many different task-dependent processes that rely on distinct neural circuits (e.g., sensory–motor vs. sensory recognition vs. frontal decision-related circuits). Information flow within the neural networks supporting these distinct processes can be considered processing "streams." Therefore, viewed in this way, the streams are task-defined rather than stimulus feature-defined. Figure 12.5 provides a graphic representation of this distinction. The dorsal versus ventral distinction, according to this framework is an oversimplification that reflects a coarse research emphasis on broad categories of processes (e.g., object identification vs. sensory–motor integration) and that ignores any number of potential finer-grained processing streams.

This task-driven framework for understanding processing streams effectively removes "where" from consideration as a viable processing stream because "where" is not a task but a stimulus feature that can be used in the performance of many task goals (Middlebrooks, 2002). This perspective does not preclude the existence of say a cortical "spatial area" that computes spatial location information which then interacts with higher-order networks on a task-dependent basis. In other words, it is logically possible that spatial activations found in the anterior PT correspond to a "feature" processing network in the task-driven model. However, it is also logically possible that the spatial feature processing network is subcortical and the cortical activation found in "spatial" tasks reflects a task-specific network that is putting spatial information to use. This is an empirical question that needs to be addressed explicitly in future work, for example by mapping the distribution of "spatial" responses under a variety of task conditions and identifying those regions that are task-dependent versus task-independent; only the latter would be candidates for "feature processing systems."

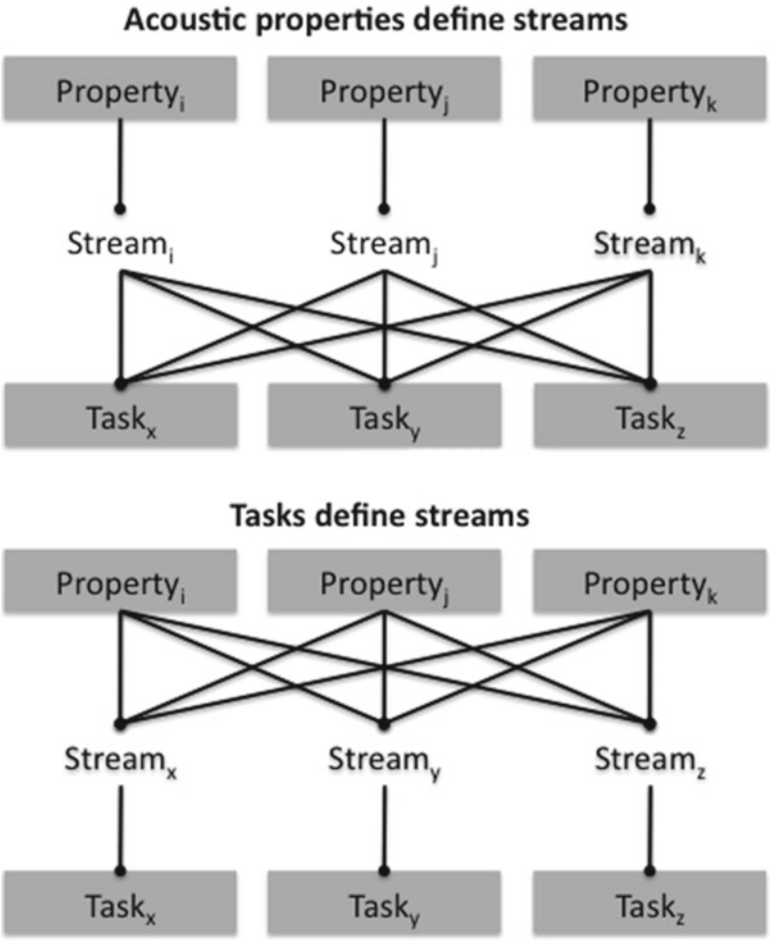


Fig. 12.5 Schematic depiction of a stimulus- versus a task-based model of sensory processing streams. See text for details

12.6 Clinical Evidence and Applications

The PT has been implicated in speech-related symptoms of at least three different disorders, conduction aphasia (noted previously), developmental stuttering, and auditory hallucinations in schizophrenia. The functional relation between the PT and these disorders are discussed in turn.

Developmental stuttering is a disorder affecting speech fluency in which sounds, syllables, or words may be repeated or prolonged during speech production. Auditory input affects fluency in people who stutter. For example, delayed auditory feedback can result in a paradoxical improvement in fluency (Martin & Haroldson, 1979; Stuart et al., 2008). This paradoxical delayed auditory feedback effect is correlated with planum temporale asymmetry. In one study, stutterers who show the paradoxical delayed auditory feedback effect also had a reversed PT asymmetry (right > left) (Foundas et al., 2004) (recall that PT asymmetry is primarily driven by area Tpt in the posterior PT). And as noted previously, altered auditory feedback modulates activity in the PTR (Fig. 12.3D). Thus, an association exists between the posterior PT, sensory–motor integration, and people who stutter suggesting that Spt (dys)function is involved in this clinical population. It has been suggested that stuttering is caused by dysfunction of internal models involved in motor control of speech, which may result in an over-reliance on sensory feedback that is substantially delayed relative to internal control mechanisms (Max et al., 2004).. The work reviewed in this chapter suggests that the posterior PT, area Spt in particular, will be a profitable focus of investigation in this respect.

A prominent positive symptom of schizophrenia is auditory hallucinations, typically involving perceived "voices." It has (recently) been suggested that this symptom results from imprecise motor-to-sensory corollary discharges (Heinks-Maldonado et al., 2007). Self-generated actions have sensory consequences; for example, moving one's eyes results in the movement of the visual field across the retina. Yet we do not perceive this sweep across the retina as motion but rather perceive a stable external environment. This is achieved by sending a corollary discharge (forward model) of the motor command to sensory areas, which can be compared against the incoming sensory information to effectively cancel the sensory consequences of self-generated actions. A similar mechanism appears to hold for speech as well, as indicated by the observation that the auditory response to speech is suppressed when speech is self generated (Paus et al., 1996; Heinks-Maldonado et al., 2007). If corollary discharges associated with speech acts (1) are used to distinguish self- from externally generated speech, and (2) if this system is imprecise in schizophrenia, self-generated speech (perhaps even subvocal speech) may be perceived as externally generated, that is, hallucinations. Consistent with this hypothesis, hallucinating patients do not show the normal suppression of auditory response to self-generated speech and the degree of abnormality correlated both with severity of hallucinations and misattributions of self-generated speech (Heinks-Maldonado et al., 2007). Schizophrenics also have anatomical abnormalities of the PT, particularly in the upper cortical layers (I–III, the corticocortical layers) of the caudal PT (~Tpt) in the left hemisphere, which show a reduced fractional volume relative to controls (Smiley et al., 2009). Thus, in schizophrenia the nature of the behavioral and physiological effects (implicating sensory–motor integration,) the location of anatomical abnormalities (left posterior PT), and the level of cortical processing implicated (corticocortical) are all consistent with dysfunction involving area Spt. As with stuttering, a research emphasis on this functional circuit is warranted in understanding aspects of schizophrenia.

One would not have expected a connection between disorders as apparently varied as conduction aphasia, stuttering, and schizophrenia, yet they all seem to involve, in part, dysfunction of the same region and functional circuit. A closer look at these syndromes reveals other similarities. For example, all three conditions show atypical responses to delayed auditory feedback. Fluency of speech in both people who stutter and conduction aphasics is not negatively affected by delayed auditory feedback and may show paradoxical improvement (Boller et al., 1978; Martin & Haroldson, 1979; Stuart et al., 2008), whereas in schizophrenia delayed auditory feedback induces the reverse effect: greater than normal speech dysfluency (Goldberg et al., 1997). Further, both stuttering and schizophrenia appear to be associated with dopamine abnormalities: dopamine antagonists such as risperidone and olanzapine (atypical antipsychotics commonly used to treat schizophrenia) have recently been shown to reduce stuttering (Maguire et al., 2004). It is unclear how dysfunction of what appears to be the same circuit can result in the range of speech/hearing symptoms found in conduction aphasia, stuttering, and schizophrenia. Rather than a problem, however, having a variety of breakdown scenarios may prove to be particularly instructive in working out the details of the circuit.

12.7 Conclusions and Remaining Questions

Neuroanatomical and neurophysiological evidence indicates that the planum temporale is functionally subdivided into (1) an anterior sector that is part of auditory cortex proper and that supports spatial-related but not necessarily spatial-specific functions (such as stream segregation), and (2) a posterior sector that is not part of auditory cortex and which supports sensory–motor integration for vocal tract actions. These functions are likely not restricted to the PT but extend beyond its anatomical boundary to involve cortex extending into the parietal operculum and the superior temporal sulcus. It is also likely that this broader PT region contains further functional subdivisions. For example, the existence of sensory–motor integration processes at both the segmental (individual phonemes) and suprasegmental levels (e.g., pitch and sequences of sounds) was mentioned earlier. There may be distinct, parallel circuits involved in sensory-motor integration at these different levels. Similarly, the cytoarchitectonic subdivisions of the anterior PTR (PaAi, PaAe, PaA c/d) may underlie functional subdivisions between these auditory areas. These issues will require further investigation using within subjects designs and high spatial resolution approaches.

A major functional component of the PTR is sensory–motor integration, particularly for vocal tract actions. Although this circuit has been characterized as the dorsal auditory stream, it seems to be neither purely auditory (Hickok et al., 2009; Okada & Hickok, 2009) nor the only possible dorsal target for auditory information, which also interacts with posterior parietal areas controlling a range of movement systems (Grunewald et al., 1999; Lewis & Van Essen, 2000; Britten, 2008). In light of these observations, we have proposed a refined conceptualization of sensory processing "streams" whereby a stream is defined not by the kinds of computations that are performed within a sensory modality (e.g., pitch vs. location) but by the kinds of task-determined supramodal systems with which a sensory system must interact (e.g., conceptual semantic vs. motor control). On this view, processing streams are not part of a single sensory modality (Pa & Hickok, 2008), rendering terms such as "*auditory* dorsal stream" or "*visual* dorsal stream" outdated at best and misleading at worst. Further, this view moves beyond simple dichotomies, which increasingly fall short in explaining the range of empirical observations (Rossetti et al., 2003; Pisella et al., 2009), and affords the possibility that the same sensory information (e.g., location) can enter into multiple higher-order processing streams depending on how that information is put to use.

Acknowledgments The authors thank Troy Hackett for his assistance on anatomical issues. This work was supported by NIH R01 DC009659.

References

Altmann, C. F., Bledowski, C., Wibral, M., & Kaiser, J. (2007). Processing of location and pattern changes of natural sounds in the human auditory cortex. *NeuroImage*, 35, 1192–1200.

Andersen, R. (1997). Multimodal integration for the representation of space in the posterior parietal cortex. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 352, 1421–1428.

Anderson, J. M., Gilmore, R., Roper, S., Crosson, B., Bauer, R. M., Nadeau, S., et al. (1999). Conduction aphasia and the arcuate fasciculus: A reexamination of the Wernicke-Geschwind model. *Brain and Language*, 70, 1–12.

Baldo, J. V., Klostermann, E. C., & Dronkers, N. F. (2008). It's either a cook or a baker: patients with conduction aphasia get the gist but lose the trace. *Brain and Language*, 105(2), 134–140.

Barrett, D. J., & Hall, D. A. (2006). Response preferences for "what" and "where" in human non-primary auditorysss cortex. *NeuroImage*, 32(2), 968–977.

Baumgart, F., Gaschler-Markefski, B., Woldorff, M. G., Heinze, H. J., & Scheich, H. (1999). A movement-sensitive area in auditory cortex. *Nature*, 400(6746), 724–726.

Benson, D. F., Sheremata, W. A., Bouchard, R., Segarra, J. M., Price, D., & Geschwind, N. (1973). Conduction aphasia: A clincopathological study. *Archives of Neurology*, 28, 339–346.

Binder, J. T., Frost, J. A., Hammeke, T. A., Rao, S. M., & Cox, R. W. (1996). Function of the left planum temporale in auditory and linguistic processing. *Brain*, 119, 1239–1247.

Boller, F., Vrtunski, P. B., Kim, Y., & Mack, J. L. (1978). Delayed auditory feedback and aphasia. *Cortex*, 14(2), 212–226.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

Britten, K. H. (2008). Mechanisms of self-motion perception. *Annual Review of Neuroscience*, 31, 389–410.

Buchsbaum, B., Hickok, G., & Humphries, C. (2001). Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cognitive Science*, 25, 663–678.

Buchsbaum, B. R., Baldo, J., Okada, K., Berman, K. F., Dronkers, N., D'Esposito, M., & Hickok, G. (2011). Conduction aphasia, sensory-motor integration, and phonological short-term memory - An aggregate analysis of lesion and fMRI data. *Brain and Language*, 119(3), 119–28.

Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *Journal of the Acoustical Society of America*, 103(6), 3153–3161.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., et al. (1997). Activation of auditory cortex during silent lipreading. *Science*, 276(5312), 593–596.

Christoffels, I. K., Formisano, E., & Schiller, N. O. (2007). Neural correlates of verbal feedback processing: An fMRI study employing overt speech. *Human Brain Mapping*, 28(9), 868–879.

Colby, C. L., & Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annual Review of Neuroscience*, 22, 319–349.

Dorsaint-Pierre, R., Penhune, V. B., Watkins, K. E., Neelin, P., Lerch, J. P., Bouffard, M., & Zatorre, R. J. (2006). Asymmetries of the planum temporale and Heschl's gyrus: relationship to language lateralization. *Brain*, 129(Pt 5), 1164–1176.

Foundas, A. L., Bollich, A. M., Feldman, J., Corey, D. M., Hurley, M., Lemen, L. C., & Heilman, K. M. (2004). Aberrant auditory processing and atypical planum temporale in developmental stuttering. *Neurology*, 63(9), 1640–1646.

Galaburda, A., & Sanides, F. (1980). Cytoarchitectonic organization of the human auditory cortex. *Journal of Comparative Neurology*, 190, 597–610.

Galaburda, A. M. (1982). Histology, architectonics, and asymmetry of language areas. In M. A. Arbib, D. Caplan & J. C. Marshall (Eds.), *Neural models of language processes* (pp. 435–445). San Diego: Academic Press.

Galaburda, A. M., Sanides, F., & Geschwind, N. (1978). Human brain. Cytoarchitectonic left-right asymmetries in the temporal speech region. *Archives of Neurology*, 35(12), 812–817.

Gannon, P. J., Holloway, R. L., Broadfield, D. C., & Braun, A. R. (1998). Asymmetry of the chimpanzee planum temporale: Humanlike pattern of Wernicke's brain language area homolog. *Science*, 279, 220–222.

Geschwind, N., & Levitsky, W. (1968). Human brain: Left-right asymmetries in temporal speech region. *Science*, 161, 186–187.

Goldberg, T. E., Gold, J. M., Coppola, R., & Weinberger, D. R. (1997). Unnatural practices, unspeakable actions: a study of delayed auditory feedback in schizophrenia. *American Journal of Psychiatry*, 154(6), 858–860.

Golfinopoulos, E., Tourville, J. A., & Guenther, F. H. (2010). The integration of large-scale neural network modeling and functional brain imaging in speech motor control. *NeuroImage*, 52(3), 862–874.

Goodglass, H. (1992). Diagnosis of conduction aphasia. In S. E. Kohn (Ed.), *Conduction aphasia* (pp. 39–49). Hillsdale, NJ: Lawrence Erlbaum Associates.

Griffiths, T. D., & Warren, J. D. (2002). The planum temporale as a computational hub. *Trends in Neuroscience*, 25(7), 348–353.

Grunewald, A., Linden, J. F., & Andersen, R. A. (1999). Responses to auditory stimuli in macaque lateral intraparietal area. I. Effects of training. *Journal of Neurophysiology*, 82(1), 330–342.

Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review*, 105, 611–633.

Hackett, T. A., De La Mothe, L. A., Ulbert, I., Karmos, G., Smiley, J., & Schroeder, C. E. (2007). Multisensory convergence in auditory cortex, II. Thalamocortical connections of the caudal superior temporal plane. *Journal of Computational Neurology*, 502(6), 924–952.

Heinks-Maldonado, T. H., Mathalon, D. H., Houde, J. F., Gray, M., Faustman, W. O., & Ford, J. M. (2007). Relationship of imprecise corollary discharge in schizophrenia to auditory hallucinations. *Archives of General Psychiatry*, 64(3), 286–296.

Hickok, G. (2009). The functional neuroanatomy of language. *Physics of Life Reviews*, 6, 121–143.

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, 4, 131–138.

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, 92, 67–99.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393–402.

Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory-motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. *Journal of Cognitive Neuroscience*, 15, 673–682.

Hickok, G., Okada, K., & Serences, J. T. (2009). Area Spt in the human planum temporale supports sensory-motor integration for speech processing. *Journal of Neurophysiology*, 101(5), 2725–2732.

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron*, 69(3), 407–422.

Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, 279, 1213–1216.

Isenberg, A. L., Vaden, K. I., Jr., Saberi, K., Muftuler, L. T., & Hickok, G. (2011). Functionally distinct regions for spatial processing and sensory-motor integration in the planum temporale. *Human Brain Mapping*. doi: 10.1002/hbm.21373.

Jacquemot, C., Dupoux, E., & Bachoud-Levi, A. C. (2007). Breaking the mirror: Asymmetrical disconnection between the phonological input and output codes. *Cognitive Neuropsychology*, 24(1), 3–22.

Lewis, J. W., & Van Essen, D. C. (2000). Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. *Journal of Comparative Neurology*, 428(1), 112–137.

Maguire, G. A., Yu, B. P., Franklin, D. L., & Riley, G. D. (2004). Alleviating stuttering with pharmacological interventions. *Expert Opinion in Pharmacotherapy*, 5(7), 1565–1571.

Martin, R., & Haroldson, S. K. (1979). Effects of five experimental treatments of stuttering. *Journal of Speech and Hearing Research*, 22(1), 132–146.

Max, L., Guenther, F. H., Gracco, V. L., Ghosh, S. S., & Wallace, M. E. (2004). Unstable or insufficiently activated internal models and feedback-biased motor control as sournces of dysfluency: A theoretical model of stuttering. *Contemporary Issue in Communication Science and Disorders*, 31, 105–122.

Middlebrooks, J. C. (2002). Auditory space processing: Here, there or everywhere? *Nature Neuroscience*, 5(9), 824–826.

Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action*. Oxford: Oxford University Press.

Okada, K., & Hickok, G. (2009). Two cortical mechanisms support the integration of visual and auditory speech: A hypothesis and preliminary data. *Neuroscience Letters*, 452(3), 219–223.

Overath, T., Cusack, R., Kumar, S., von Kriegstein, K., Warren, J. D., Grube, M., et al. (2007). An information theoretic characterisation of auditory encoding. *PLoS Biology*, 5(11), e288.

Pa, J., & Hickok, G. (2008). A parietal-temporal sensory-motor integration area for the human vocal tract: Evidence from an fMRI study of skilled musicians. *Neuropsychologia*, 46, 362–368.

Paus, T., Perry, D. W., Zatorre, R. J., Worsley, K. J., & Evans, A. C. (1996). Modulation of cerebral blood flow in the human auditory cortex during speech: Role of motor-to-sensory discharges. *European Journal of Neuroscience*, 8, 2236–2246.

Pisella, L., Sergio, L., Blangero, A., Torchin, H., Vighetto, A., & Rossetti, Y. (2009). Optic ataxia and the function of the dorsal stream: Contributions to perception and action. *Neuropsychologia*, 47(14), 3033–3044.

Rauschecker, J. P. (1998). Cortical processing of complex sounds. *Current Opinion in Neurobiology*, 8(4), 516–521.

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6), 718–724.

Rossetti, Y., Pisella, L., & Vighetto, A. (2003). Optic ataxia revisited: Visually guided action versus immediate visuomotor control. *Experimental Brain Research*, 153(2), 171–179.

Schlaug, G., Jancke, L., Huang, Y., & Steinmetz, H. (1995). In vivo evidence of structural brain asymmetry in musicians. *Science*, 267, 699–701.

Smiley, J. F., Hackett, T. A., Ulbert, I., Karmas, G., Lakatos, P., Javitt, D. C., & Schroeder, C. E. (2007). Multisensory convergence in auditory cortex, I. Cortical connections of the caudal superior temporal plane in macaque monkeys. *Journal of Computational Neurology*, 502(6), 894–923.

Smiley, J. F., Rosoklija, G., Mancevski, B., Mann, J. J., Dwork, A. J., & Javitt, D. C. (2009). Altered volume and hemispheric asymmetry of the superficial cortical layers in the schizophrenia planum temporale. *European Journal of Neuroscience*, 30(3), 449–463.

Smith, K. R., Okada, K., Saberi, K., & Hickok, G. (2004). Human cortical motion areas are not motion selective. *NeuroReport*, 9, 1523–1526.

Smith, K. R., Saberi, K., & Hickok, G. (2007). An event-related fMRI study of auditory motion perception: No evidence for a specialized cortical system. *Brain Research*, 1150, 94–99.

Smith, K. R., Hsieh, I. H., Saberi, K., & Hickok, G. (2009). Auditory spatial and object processing in the human planum temporale: No evidence for selectivity. *Journal of Cognitive Neuroscience*.

Stuart, A., Kalinowski, J., Rastatter, M. P., & Lynch, K. (2002). Effect of delayed auditory feedback on normal speakers at two speech rates. *Journal of the Acoustic Society of America*, 111(5 Pt 1), 2237–2241.

Stuart, A., Frazier, C. L., Kalinowski, J., & Vos, P. W. (2008). The effect of frequency altered feedback on stuttering duration and type. *Journal of Speech Language and Hearing Research*, 51(4), 889–897.

Sweet, R. A., Dorph-Petersen, K. A., & Lewis, D. A. (2005). Mapping auditory core, lateral belt, and parabelt cortices in the human superior temporal gyrus. *Journal of Computational Neurology*, 491(3), 270–289.

Takaso, H., Eisner, F., Wise, R. J., & Scott, S. K. (2010). The effect of delayed auditory feedback on activity in the temporal lobe while speaking: A positron emission tomography study. *Journal of Speech Language and Hearing Research*, 53(2), 226–236.

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage*, 39(3), 1429–1443.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale & R. J. W. Mansfield (Eds)., *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the USA*, 102(4), 1181–1186.

Warren, J. D., & Griffiths, T. D. (2003). Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. *The Journal of Neuroscience*, 23, 5799–5804.

Warren, J. D., Zielinski, B. A., Green, G. G., Rauschecker, J. P., & Griffiths, T. D. (2002). Perception of sound-source motion by the human brain. *Neuron*, 34(1), 139–148.

Warren, J. E., Wise, R. J., & Warren, J. D. (2005). Sounds do-able: Auditory-motor transformations and the posterior temporal plane. *Trends in Neuroscience*, 28(12), 636–643.

Wise, R. J. S., Scott, S. K., Blank, S. C., Mummery, C. J., Murphy, K., & Warburton, E. A. (2001). Separate neural sub-systems within "Wernicke's area." *Brain*, 124, 83–95.

Yates, A. J. (1963). Delayed auditory feedback. *Psychological Bulletin*, 60, 213–251.

Zatorre, R. J., Bouffard, M., Ahad, P., & Belin, P. (2002). Where is where' in the human auditory cortex? *Nature Neuroscience*, 5(9), 905–909.

Zheng, Z. Z., Munhall, K. G., & Johnsrude, I. S. (2009). Functional overlap between regions involved in speech perception and in monitoring one's own voice during speech production. *Journal of Cognitive Neuroscience*, 22(8), 1770–1781.

Chapter 13
Toward a Theory of Information Processing in Auditory Cortex

Peter Cariani and Christophe Micheyl

13.1 Introduction

The primary goal of auditory science is to provide a full account of how humans and animals hear sounds of all kinds: the sounds of everyday life, environmental sounds, speech, and music. A comprehensive, neurally grounded theory of hearing is needed that explains precisely how we hear what we hear. This chapter discusses cortical function in the context of such a theory. The first half of the chapter (Sections 13.2 and 13.3) outlines the basic structure of hearing (what is to be explained) and the various aspects of neural information processing that are needed for adequate explanations of auditory function (the terms of the explanations). The second half (Section 13.4) lists some of the fundamental outstanding experimental and theoretical problems that need to be solved.

The goal of auditory theory is to understand how the auditory system works as an informational system. Once such a theory is finally formulated, such that a firm understanding of the codes, computations, and their neuronal substrates is achieved, then more effective therapeutic strategies for restoring auditory functions lost to disease can be devised, and artificial devices that expand the sound analysis capabilities of our own natural auditory systems can be designed and built.

What would a complete theory of audition entail? Such a theory would identify and account for the perceptual and cognitive *informational functions* that the auditory system carries out (Fig. 13.1), as well as the structure of subjective auditory

P. Cariani (✉)
Department of Otology and Laryngology, Harvard Medical School,
629 Watertown Street, Newton, MA 02460, USA
e-mail: cariani@mac.com

C. Micheyl
Auditory Perception and Cognition Laboratory, Department of Psychology,
University of Minnesota, N628 Elliot Hall, 75 East River Parkway,
Minneapolis, MN 55455, USA
e-mail: cmicheyl@umn.edu

D. Poeppel et al. (eds.), *The Human Auditory Cortex*, Springer Handbook
of Auditory Research 43, DOI 10.1007/978-1-4614-2314-0_13,

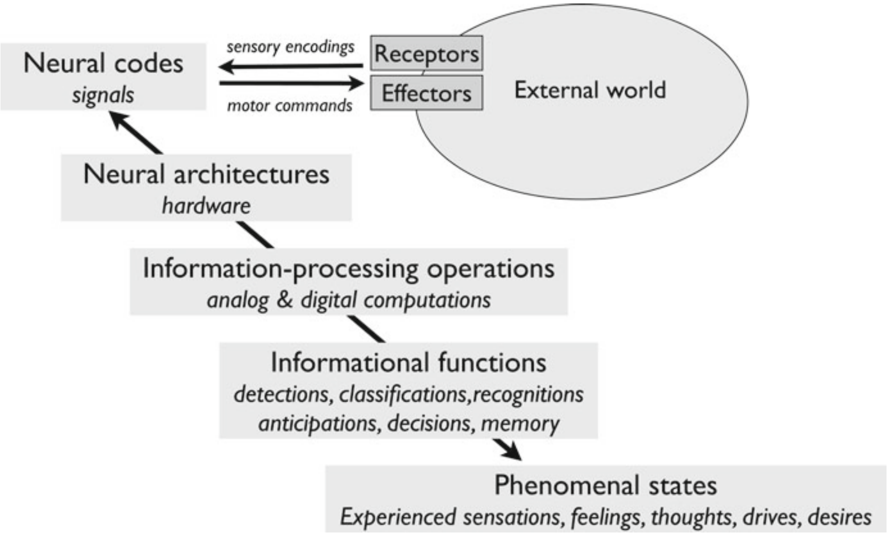

Neural codes
signals
sensory encodings
motor commands
Receptors
Effectors
External world
Neural architectures
hardware
Information-processing operations
analog & digital computations
Informational functions
detections, classifications,recognitions
anticipations, decisions, memory
Phenomenal states
Experienced sensations, feelings, thoughts, drives, desires

Fig. 13.1 Complementary aspects of neural information processing that are essential for understanding how the auditory system works. Auditory theory seeks to explain auditory functions and experiences in terms of neural codes, architectures, and information-processing operations. These aspects can be compared with Marr's (1982) division into computational, algorithmic, and implementational levels of description

experience (*phenomenal states*). Accounting for auditory functions in neural terms involves identification of neural codes, neuronal architectures, and neurocomputational operations. *Neural codes* are the basic signals of the system that support systematic internal representations of sound attributes. *Neuronal architectures* are the neuronal hardware substrates that implement neural signal processing. The rubric of architectures includes the composition and organization of neural elements, from molecular, cellular, and anatomical structures and functions to patterns of interneural connections. *Information-processing operations* on internal representations carry out transformations, analyses, and decisions that ultimately steer and switch behavior. These aspects of the auditory system tell us how the system must be organized so as to achieve its functions at any given time. However, an even fuller account of audition also includes a historical understanding of how the system came to be. This includes the ontogenetic, developmental processes that construct and modify the auditory system over the life of the organism and the phylogenetic, evolutionary processes that have shaped its structures and functions, from its first appearance in ancient animals to its present-day form.

As with many other structure–function relations in biology, each aspect is complementary to the rest. A given function, being a set of relations, does not uniquely determine the underlying structures and mechanisms that implement it. For example, auditory functions can potentially be carried out in different ways, using different neural codes, operations, and architectures that utilize different biophysical mechanisms. Conversely, although underlying molecular, cellular, and network structures and biophysics constrain which codes, operations, and architectures are possible, complete knowledge of underlying neuronal substrates does not necessarily directly lead to better understanding of the codes, operations, and architectures that realize auditory functions.

Taken together, these complementary aspects of neural information processing yield a comprehensive, systematic account of how the system works: what functions it realizes, what kinds of internal signals and representations it uses, what processing operations are carried out, as well as the neural network architectures and properties of the neuronal system elements that are essential to its functioning. Such an account would constitute a full mechanistic and neurocomputational model of auditory structures, mechanisms, and functions.

Even so sweeping an account, however, would omit the phenomenal, subjective aspect of hearing that we as conscious subjects experience directly when we hear a sound. In the last two decades, there has been a resurgence of interest and investigation in the neuronal correlates of subjective experiences (Koch, 2004). The neural correlates of consciousness (NCCs) involve the neuronal requisites for having conscious awareness, while the neural correlates of the contents of consciousness (NCCCs) involve the specific neural basis for particular experiences.

Although most empirical investigation and theory-building in this emerging field of consciousness studies has involved the visual system, involving phenomena such as visual masking, diverted attention, blindsight, and visual neglect syndromes (Pollen, 2008), close auditory analogues of all these phenomena exist. Thus, there is no reason why the neural basis of auditory experience cannot furnish comparable insights into the neural basis for conscious awareness. What are the minimal, requisite neuronal conditions for a sound to be consciously experienced (auditory NCCs), and what patterns of neuronal activity correspond to which changes in auditory experience (auditory NCCCs)? These are fundamental questions that are also practically relevant for understanding the neurophenomenological basis of auditory hallucinations and tinnitus.

A full theory of audition thus should explain the relations between sounds, neuronal responses, auditory functions, and auditory experience (Fig. 13.2). Psychoacoustics and psychology address questions concerning relations between acoustic stimuli and auditory perception and cognition. Neurophysiology and computational neuroscience seek to characterize neuronal responses to sounds (system identification) and to identify correspondences between neuronal responses and auditory functions (neural coding) such that underlying information processing principles that are utilized by the auditory system can be identified (reverse engineering). Lastly, one can point to a future field of auditory neurophenomenology that would elucidate the structure of subjective auditory experience and formulate neurophenomenal bridge laws that predict the dimensional structure and contents of experience (NCCs and NCCCs) from neural activity patterns.

This chapter begins with the basic auditory functions that auditory theory seeks to explain (Section 13.2) followed by discussion of the nature of such an explanation in terms of the codes, computational operations, and architectures that might be

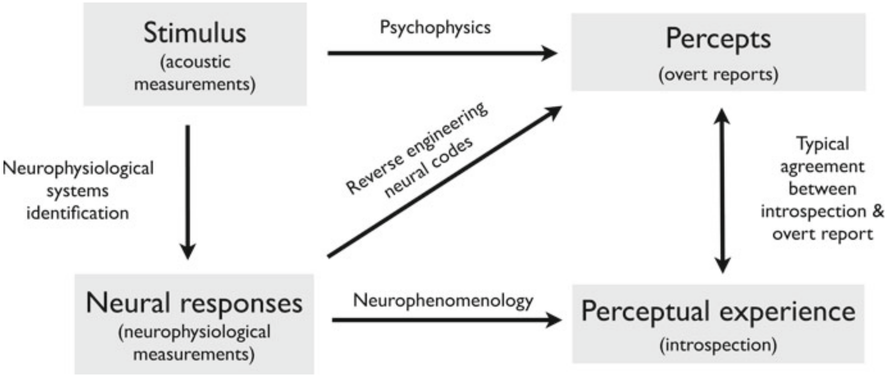

Stimulus
(acoustic measurements)
Psychophysics
Percepts
(overt reports)
Neurophysiological systems identification
Reverse engineering neural codes
Typical agreement between introspection & overt report
Neural responses
(neurophysiological measurements)
Neurophenomenology
Perceptual experience
(introspection)

Fig. 13.2 Stimulus–response relations and neuropsychological states. Auditory psychophysics involves modeling of stimulus–percept relations. Neurophysiological systems identification involves prediction of the behavior of neural elements and networks as a function of the acoustic stimulus. The neural coding problem involves identifying, through reverse engineering, which aspects of neural response are causally related to informational (perceptual) functions. Neurophenomenology involves identifying which aspects of neural activity are necessary and/or sufficient to produce particular experiential states. Public and private measurements associated with acoustic, neural, behavioral, and phenomenal states are listed in parentheses

involved (Section 13.3). Succeeding sections then take up outstanding problems and fundamental questions about how the auditory system works and the role auditory cortex might play in that process. These questions involve identifying the neural codes operant at the cortical level and understanding the neurocomputational basis for invariances and invariant transformations of auditory percepts as well as their precision and robustness. This discussion seeks to present the fundamental scientific questions that auditory theory faces in as comprehensive and clear manner as possible, to provoke deeper thinking for more comprehensive theory-building that can guide experimental investigations. Unlike simple reports of empirical data, which usually stand on their own, these ideas are necessarily exploratory and speculative, and are meant to widen rather than narrow the realm of possible mechanisms for consideration.

13.2 What a Neurocomputational Theory of Auditory Cortex Seeks to Explain

13.2.1 General Auditory Functions

The primary goal of a theory of audition is to explain auditory function. This first entails a broad, ecological account of how the auditory system enhances the survival and reproductive fitness of the organism and its lineage as well as specific accounts

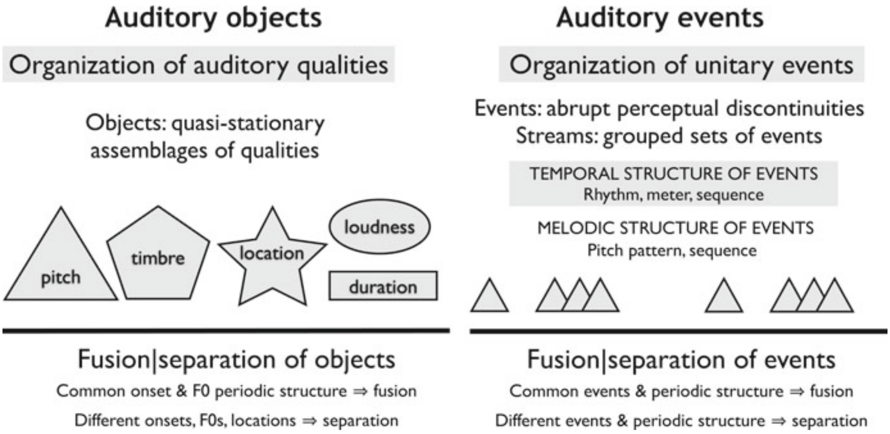

Auditory objects
Organization of auditory qualities
Objects: quasi-stationary assemblages of qualities
pitch
timbre
location
loudness
duration
Fusion|separation of objects
Common onset & F0 periodic structure ⇒ fusion
Different onsets, F0s, locations ⇒ separation
Auditory events
Organization of unitary events
Events: abrupt perceptual discontinuities
Streams: grouped sets of events
TEMPORAL STRUCTURE OF EVENTS
Rhythm, meter, sequence
MELODIC STRUCTURE OF EVENTS
Pitch pattern, sequence
Fusion|separation of events
Common events & periodic structure ⇒ fusion
Different events & periodic structure ⇒ separation

Fig. 13.3 Perceptual organization of auditory qualities and events. (Left) Grouping of auditory qualities in objects and major factors that govern the fusion or separation of objects. (Right) Grouping of events into streams and major factors that govern the fusion or separation of streams

of particular perceptual and cognitive functions that the auditory system realizes, such as detection, discrimination, object and stream formation and separation, analysis, and recognition of sounds. These specific functions include representation and discrimination of different physical aspects of sounds, such as periodicity, spectrum, intensity, duration, and their spatial relationships to listeners. They also include auditory functionalities related to inferences about the auditory scene (how many independent sound sources and the properties of the individual sounds) as well as learned past experiences of sound that have been retained in memory: familiarity, hedonic preference (euphonious vs. unpleasant attributes), category (e.g., phonetic class, or for absolute pitch possessors, pitch class), sound source identity (what speaker, musical instrument, animal, or natural process has produced a given sound?). These functions all provide critical information about environmental sound sources that facilitate their recognition. Auditory functions support more appropriate and effective responses to environmental events.

A comprehensive theory of auditory cortical function should explain how perception of different qualities of sound (pitch, timbre, loudness, duration, location) is subserved by neuronal cortical representations and mechanisms. How do human and animal listeners discriminate fine differences to differentiate sounds and generalize similarities to recognize categories of sound? How does auditory theory account for the dimensional structure of auditory perception—the nearly complete independence of different perceptual attributes? Such a theory should also explain why the auditory scene has the organization that it does, such that separate objects and events are distinguished and their respective attributes grouped together. The auditory scene at any given time contains a set of quasi-stable objects that have an organization in terms of their respective attributes (Fig. 13.3). Analogously, there is a larger organization of the auditory scene in time, in which longer temporal patterns of related events cohere into unified patterns (e.g., rhythmic and melodic sequences), with their separation and fusion into streams (stream segregations and fusions).

Basic auditory attributes are discussed first, followed by their organization in terms of object representations, and the temporal organization of auditory events and objects into streams.

13.2.2 Organization of the Auditory Scene

Perception has a strong dimensional structure, both at the level of different sensory modalities and kinds of distinctions within each modality (Boring, 1942). In humans, major sensory modalities include audition, vision, olfaction, gustation, balance and orientation, proprioception, pain, thermoreception, flutter-vibration, pressure, as well as a host of various interoreceptors. In turn, audition has a strong dimensional structure that is reflected in basic attributes of sound perception: loudness, pitch, timbre, duration, and apparent location. There is also a higher-level organization of these primitive auditory qualities in which some sets of attributes are grouped together in objects (Fig. 13.3). At any present moment, auditory experience consists of a temporal succession of perceived objects and events. Auditory scene analysis involves accounting for how the auditory system organizes incoming sound patterns into auditory objects and events, each with its own set of associated perceptual attributes (Handel, 1989; Bregman, 1990). The general principles that underlie auditory perceptual organization were originally investigated by Gestaltist psychologists (e.g., Köhler, 1947), and later came to be popularly known as the "cocktail party problem" (Cherry, 1966). Such organizing principles play a particularly important role in structuring musical experiences and expectations (Handel, 1989; Bregman, 1990; Snyder, 2000).

An auditory *object* is a representational structure that is a collection of stable, quasi-invariant perceptual properties (attributes, features) that persist together over time as a unitary entity. In contrast, an auditory *event* is a representation of a salient change in the perceived auditory scene from one moment to the next that also has a set of properties (attributes, features, contrasts) associated with it. Auditory objects and events are mental constructs that may or may not correspond directly to particular, identifiable physical objects, sound sources, or sonic events in the external world. Such constructs are produced by internal representational structures that are implemented through organized patterns of neuronal activity.

A prime example of an auditory object is the perception of a single note played by a musical instrument (the temporal stability of the note being emphasized), whereas an example of an auditory event could involve perception of different, temporal aspects of the same note (the temporal succession of the note-object's appearance and disappearance being emphasized). One can also consider the onset and offset of the note as separate auditory events in and of themselves. Auditory events are mental constructs typically produced by temporal acoustic contrasts that distinguish subsequent from preceding sound patterns. The perception of a discrete event is itself the result of a temporal auditory grouping process. In most music, the onset of each note is an individual auditory event. Event boundaries in speech can be more fluid, where individual phonetic elements or rapid sequences of "chunked" elements can constitute separate, unified events.

13.2.3 Basic Auditory Qualities

Perceptual auditory qualities associated with each object or event can be grouped under more general categories of pitch, timbre, location, loudness, and duration (Fig. 13.3, left). In turn, each category can include one or more related perceptual dimensions. Some attributes, such as loudness and duration, are one dimensional, such that their qualities can be ordered in a monotonic, linear order. Location here includes several spatial qualities of sounds that include the apparent direction, extent, and range of the sound image in auditory space. Other qualities, such as pitch and timbre, are multidimensional, having several related, but distinct aspects that may be related to different underlying neural representations.

This high level division of auditory qualities is most obvious in musical contexts. Operationally, pitch is that auditory quality that covaries with the frequency of pure tones. Instruments are generally recognizable by their distinctive timbres, irrespective of the pitch, loudness, and duration of the notes played and of their position in auditory space. Instruments that produce extremely different timbres can play notes with the same recognizable pitches, and the pitches and timbres that they produce are generally highly invariant with respect to sound intensity. Although extremely short sound durations can alter perception of pitch, timbre, and loudness, these qualities are highly invariant for longer durations (>50 ms). Whereas tonal music typically involves melodic sequences of auditory events (notes) having varying pitches with relatively fixed timbres, speech communication systems typically use sequences of distinctive auditory events (phonetic elements) that have different timbres.

13.2.3.1 Loudness, Duration and Spatial Attributes

Each auditory quality has primary acoustical correlates. Loudness covaries monotonically with intensity, perceived duration with duration. Directionality in the horizontal, azimuthal plane depends on interaural time-of-arrival and sound pressure level differences at the two ears. Perceived elevation depends on high-frequency spectral notches that are characteristic of a listener's pinnae. Many physical correlates of attributes associated with spatial hearing, such as apparent distance, source width/extent, and enclosure size, are joint properties of sound sources and their reflections (Ando & Cariani, 2009).

13.2.3.2 Pitch

Pitch is somewhat more complex in structure. Operationally, the pitch of a sound is the perceptual quality that covaries with the frequency of pure tones, that is, it is defined as the frequency of a pure tone to which the perceived pitch quality of given stimulus is matched. Pitch is related to the dominant periodicity of sounds, the repetition rate of a sound pattern. For pure tones this repetition rate is simply the tone's frequency; for complex tones this repetition rate is the fundamental frequency (de Cheveigné, 2005).

Arguably there are two closely related percepts that are associated with pitch, here called "spectral pitch" and "periodicity pitch." Differences in some of their properties suggest that these percepts may be subserved by qualitatively different neural representations. Pitch height is related to the perception of the absolute "lowness or highness" of a spectral pitch, and is monotonically related to the absolute frequencies of dominant spectral components present in a sound.

Periodicity pitch is a quality related to the dominant periodicity of a sound, its fundamental frequency, irrespective of its spectral composition. It is often also called "virtual pitch," "the low pitch of complex tones," "F0 pitch," "the pitch of the (missing) fundamental," or "musical pitch," depending on context and theoretical orientation.

Low-frequency pure tones evoke both spectral and periodicity pitch. Harmonic stimuli, such as AM tones, can be constructed in which pitch height (which varies with carrier frequency) and periodicity (which varies with modulation frequency) can be independently altered. Both spectral and periodicity pitches support local judgments of whether one pitch is lower or higher than another.

Pitch height and periodicity pitch appear to reflect different, independent acoustic properties, that is, absolute frequencies vs. dominant periodicities, that are likely to be mediated by different neuronal representations. At the level of the auditory nerve spectral pitch appears to be based on cochlear place of maximal excitation, whereas periodicity pitch appears to be based on spike timing information, most likely in the form of population-wide distributions of interspike interval (Cariani & Delgutte, 1996; Cariani, 1999). The cochlear place representation runs the entire frequency range of hearing (~50–20,000 Hz), whereas useable spike timing information in the auditory nerve extends only up to the limit of significant phase locking, around 4–5 kHz. At the cortical level, changes in pitch height and periodicity-pitch chroma produce neural activity in different local areas of human auditory cortex (Warren et al., 2003).

Musical tonality appears to be related to periodicity pitch and the temporal representation. Most listeners readily recognize octave similarities, musical intervals, melodies and their transpositions, as well as mistuned, and "off-key" or mistuned "sour" notes for periodicities below 4–5 kHz, that is, within the existence region for periodicity pitch. These distinctions all involve relative pitch, which involve recognitions of particular ratios of periodicities (see McDermott & Oxenham, 2008, for discussion of cortical mechanisms and Section 13.4.4 for discussion of relative pitch and melodic transposition). Although very crude melodic distinctions based on pitch height contours can be made for sequences of pure tones above this frequency limit, in the absence of periodicity pitch, recognitions of melodies and musical intervals in this register are severely degraded. Whereas periodicity pitch is highly invariant with respect to sound intensity, spectral pitch perception of high-frequency pure tones is notoriously level dependent, consistent with the notion that the two types of pitch percepts depend, respectively, on temporal and cochlear-place neural representations. Critical questions for auditory neurophysiology involve the nature of neural coding transformations that give rise to cortical representations for spectral and periodicity pitch.

Pitch strength or salience is an intensive dimension of pitch that is related to the apparent strength of a pitch. Pitch strength is analogous to color saturation in vision. Bandwidth is the primary acoustical determinant of spectral pitch strength, whereas waveform regularity/harmonicity and harmonic number are primary determinants of the strengths of periodicity pitches. These factors been used to parametrically vary periodicity pitch strength, such that one can identify cortical territories whose neural populations respond differentially to stimuli that evoke periodicity pitches (Patterson et al., 2002; Warren et al., 2003; Penagos et al., 2004).

13.2.3.3 Timbre

Timbre is perhaps the most complex perceptual category, with multiple dimensions and acoustical correlates. Timbre is most clearly illustrated in musical contexts, where it includes those qualities of sound that distinguish the same musical note (i.e., of the same pitch, duration, and loudness) played by different instruments at the same location. Some aspects of timbre depend on the gross power spectrum of stationary sounds (e.g., spectral center of gravity, spectral tilt/brightness, formant structure, and bandwidth), whereas other aspects depend on rapid modulations and fluxes in amplitude, frequency, and phase of nonstationary sounds (Handel, 1989; McAdams & Giordano, 2009).

In musical contexts, instrument resonances determine aspects of timbre related to the power spectrum (tone color, brightness), while onset and offset dynamics (attack, sustain, decay, tremolo), frequency fluxes (temporal successions of harmonics, vibrato, noise components), and phase dynamics (chorus, flanger and phaser effects) determine those aspects of timbre that are related to rapid changes in sound.

By this expansive definition of timbre, most phonetic distinctions in speech are categorical timbral distinctions. Phonetic distinctions in atonal languages either involve the gross power spectra of stationary sounds (vowels) or amplitude and frequency transients (consonants). In tonal languages, pitch levels and contours are also used alongside timbral differences to distinguish phonetic elements.

Whereas timbral space associated with stationary spectral distinctions is relatively well understood, the structure of timbral space associated with transient changes in amplitude, frequency and phase has not been systematically characterized. As with pitch, the multiplicity of timbral qualities may be associated with different aspects of neural patterns of response.

13.2.4 Formation of Auditory Objects and Events

Auditory perception consists of more than just elemental perceptual qualities—there is an organization of perceptual attributes within the "auditory scene" in which neural representations associated with particular sound components are grouped together to form representations of unitary auditory objects or events (Handel, 1989;

Bregman, 1990). The auditory scene at any particular moment can contain multiple auditory objects, with each object having its own set of associated perceptual and cognitive attributes (loudness, duration, location, pitch, timbre plus higher-order cognitive attributes) that group together (Griffiths, Micheyl, and Overath, Chapter 8).

Common harmonic structure and common onset time are the two strongest factors that respectively cause sets of frequency components to fuse into unified auditory objects and events (Fig. 13.3, bottom left). In comparison, the sound parameters associated with location, duration, and loudness produce a much weaker basis for grouping. Common harmonic structure means that the sound patterns and internal early temporal representations of groups of harmonically related frequency components form repeating, periodic patterns. Common onset of components means that the simultaneous sound patterns, whether harmonic or inharmonic, will produce the same frequency–time pattern of spike timings and firing rates. This generates a characteristic timbre for the auditory event. Thus when the same components recur at a later time, the common onset grouping has a similar timbral and pitch representation. When frequency components have neither common harmonic structure nor common onset, the mixture is inharmonic and the representations of their sound patterns do not fuse, such that multiple pitches may be heard. Thus, one can hear out the notes of several musical instruments playing at the same time, provided that the notes are not harmonically related and their onsets are not strictly synchronized. In such cases when the notes form separate objects and events, their respective qualities (pitch, timbre, loudness, duration, and location), can be heard, such that individual instruments can be identified by the timbres of the individual notes. On the other hand, some sounds (multiple broadband noises or harmonic complexes with the same fundamental frequency) fuse together into one auditory object. To the extent that different sounds have separate internal representations, they have separate sets of perceptual attributes; to the extent that they are fused together into a single representational object, their attributes become blended together. Thus the auditory scene is perceived as containing multiple auditory objects each with its own set of perceptual attributes.

13.2.5 Grouping of Events into Streams

Related auditory events in turn are grouped into distinct associated patterns called streams (Handel, 1989; Bregman, 1990). In a manner analogous to the formation and separation of objects, neural mechanisms group recurring patterns of events and sequences of similar events (e.g., common pitch, timbre, duration, location) into streams. Each stream has attributes related to the relations between the events of the stream (Fig. 13.3, bottom right). Melody is a temporal pattern of pitch changes that coheres together as a unified, recognizable sequence. If the notes are too short (<100 ms) they blend together; if they are too long (>2–3 s), only one note is retained in echoic memory and the pitch pattern is not perceived. Similarly, rhythm is a temporal pattern of the onsets and offsets of related events that are grouped together into a stream. Like melody, if the events are too closely spaced temporally, they fuse together, and the patterns of intervening time intervals are lost. If the events are too far apart, only individual events are perceived and no pattern is established. Although longer patterns of events do not cohere into palpable rhythms, repeating sequences of events on a wide range of timescales can be recognized after several cycles have been heard.

On all timescales, repeating patterns of sound and their evoked auditory events build up strong representational expectancies of their continuation (Snyder, 2000; Winkler et al., 2009). The effect is created even with arbitrary and highly artificial repeating sound patterns. Whatever the constancies or changes, repeating sequences build up strong auditory expectations of their continuation. When sequences deviate from their previous repeated patterns, a perceptually salient expectancy violation is produced.

A great deal of music relies on the creation of musical tonal and rhythmic expectancies and their violations (Handel, 1989; Bigand, 1993; Zatorre and Zarate, Chapter 10). The push–pull of violation-induced tension and the confirmation of expectancies lies at the basis of "emotion and meaning" in complex, program music that is meant for intent listening (Huron, 2006). Such expectancy violations create distinct neural signatures (Trainor & Zatorre, 2009), such as the mismatch negativity (MMN) (Näätänen et al., 2007) and other responses (Chait et al., 2007). Thus far, it appears that any perceptible deviation from an expected pattern, such as changes in loudness, pitch, timbre, duration, or event timing, creates a delayed mismatch negativity response. Such negativities are also seen for cognitive expectancies: linguistic syntactic and semantic violations as well as deviations from musical expectancies (Patel, 2008). The various negativity responses differ in their latencies relative to the time of the violation, a reflection of the neuronal populations involved and their interconnections with auditory cortical populations. These response similarities that involve different types of perceptual and cognitive attributes suggest the existence of general cortical mechanisms for the buildup of temporal pattern expectancies and their comparison with incoming temporal sequences.

13.2.6 Cognitive Dimensions

In addition to basic perceptual auditory attributes, there are also additional, cognitive aspects of sounds that are related to internal representations and past experiences. Representations associated with these cognitive dimensions may have their own grouping and comparison mechanisms that involve both auditory and nonauditory centers. These dimensions include:

- *Categorical perception.* Sound objects and events can be recognized as categorical tokens in learned symbol systems (e.g., phonetic elements in speech, pitch classes for possessors of absolute pitch).
- *Mnemonics.* A sound can be perceived as familiar or unfamiliar, depending on whether it interacts strongly with specific short-, intermediate-, or long-term

memories. Such memories also encode learned statistical dependencies between sounds.

- *Semantics.* Sounds can acquire meaning from previous experience. Sounds can be associatively linked with perceived objects and events, such that subsequent presentation and recognition engages association-related neural anticipatory-prediction mechanisms.
- *Pragmatics.* Sounds are also experienced in the context of the internal goals and drives, such that sounds can acquire relevance for goal attainment and drive reduction (e.g., a dinner bell).
- *Hedonics.* A sound can be experienced as pleasant, neutral, or unpleasant. The hedonic valence of a sound can be related to purely sensory factors (e.g., grating or very high pitched sounds) or learned associations (e.g., the sound of a bell that precedes a shock).
- *Affective dimensions.* Beyond simple pleasantness or unpleasantness, sounds and sound sequences can induce particular emotional states.

13.3 Toward a Neurocomputational Theory of Auditory Cortex: Auditory Codes, Representations, Operations, and Processing Architectures

The previous section outlined major features of auditory perception and cognition ("what we hear") that a full theory of information processing by auditory cortex should ultimately explain. Explanation of "how we hear what we hear" is framed here in terms of neural representations, operations, and processing architectures.

Identification of the neural correlates of the basic representational dimensions and organizations of auditory perception and cognition is a critical step in development of a working theory of auditory cortex. The Gestaltist concept of neuropsychological isomorphism (Köhler, 1947) is a useful working hypothesis, that is, that every dimension of auditory perceptual function and experience reflects the dimensional structure of underlying neural representations and information-processing operations on which it depends. If so, then not only do neural representations and computations explain the structure of perception and of experience, but these structures also provide strong clues as to the nature of the underlying neuronal processes.

13.3.1 Neural Codes

Neural codes involve those aspects of neuronal activity that play functional and informational roles in the nervous system, that is, they are specific patterns of activity that switch the internal states, and ultimately the overt behavior of the system (Rieke et al., 1997). Many different kinds of neural pulse codes are possible (Cariani, 1995), and whole catalogs of possible neural codes and evidence for them have been discussed and collated in the past (Perkell & Bullock, 1968). Neural coding of sensory information can be based on discharge rates, interspike interval patterns, latency patterns, interneural discharge synchronies and correlations, temporal spike-burst structure, or still more elaborate cross-neuron volley patterns. In addition, sensory coding can be based on the mass statistics of many independent neural responses (population codes) or on the joint properties of particular combinations of responses (ensemble codes).

Amidst the many ways that neural spike trains can convey sensory information are fundamentally two basic ideas: "coding-by-channel" and "coding-by-time" (Fig. 13.4, top). Channel-based codes depend on the activation of specific neural channels or of configurations of channels. Temporal codes, on the other hand, depend on the relative timings of neural discharges rather than on which particular neural channels respond how much. Temporal codes can be based on particular patterns of spikes within spike trains (temporal-pattern codes) or on the relative times-of-arrival of spikes (time-of-arrival codes).

13.3.1.1 Channel-Based Coding Schemes

Many different channel-based coding schemes are possible. Such schemes can range from simple, unidimensional representations to low-dimensional sensory maps to higher dimensional feature detectors. In simple "doorbell" or "labeled line" systems, activation (or suppression) of a given neuron signals the presence or absence of one particular property. In more multipurpose schemes, neurons are differentially tuned to particular stimulus properties, such as frequency, periodicity, intensity, duration, or external location. Profiles of average discharge rates across a population of such tuned elements then convey multidimensional information about a stimulus. When spatially organized in a systematic manner by their tunings, these elements form sensory maps, in which spatial patterns of channel activation can then represent arbitrary combinations of those stimulus properties. In lieu of coherent spatial order, tuned units can potentially convey their respective channel identities through specific connections to other neurons beyond their immediate neighborhood. More complex constellations of properties can be represented via more complex concatenations of tunings to form highly specific "feature detectors." In the absence of coherent tunings, combinations of idiosyncratic response properties can potentially support "across-neuron pattern codes" of the sort that have been proposed for the olfactory system.

Nevertheless, idiosyncratic across-neuron patterns and associative learning mechanisms present fundamental difficulties in explaining common strong perceptual equivalence classes that are shared by most humans and are largely independent of an individual's particular history. Although these various functional organizations, from labeled lines to feature detectors to across-neuron patterns, encompass widely diverse modes of neural representation, all draw on the same basic strategy of coding-by-channel. In channel-coding schemes, it is usually further assumed that distinctions between alternative signal states are encoded by different average discharge rates. The combination of channel- and rate-based coding has remained by

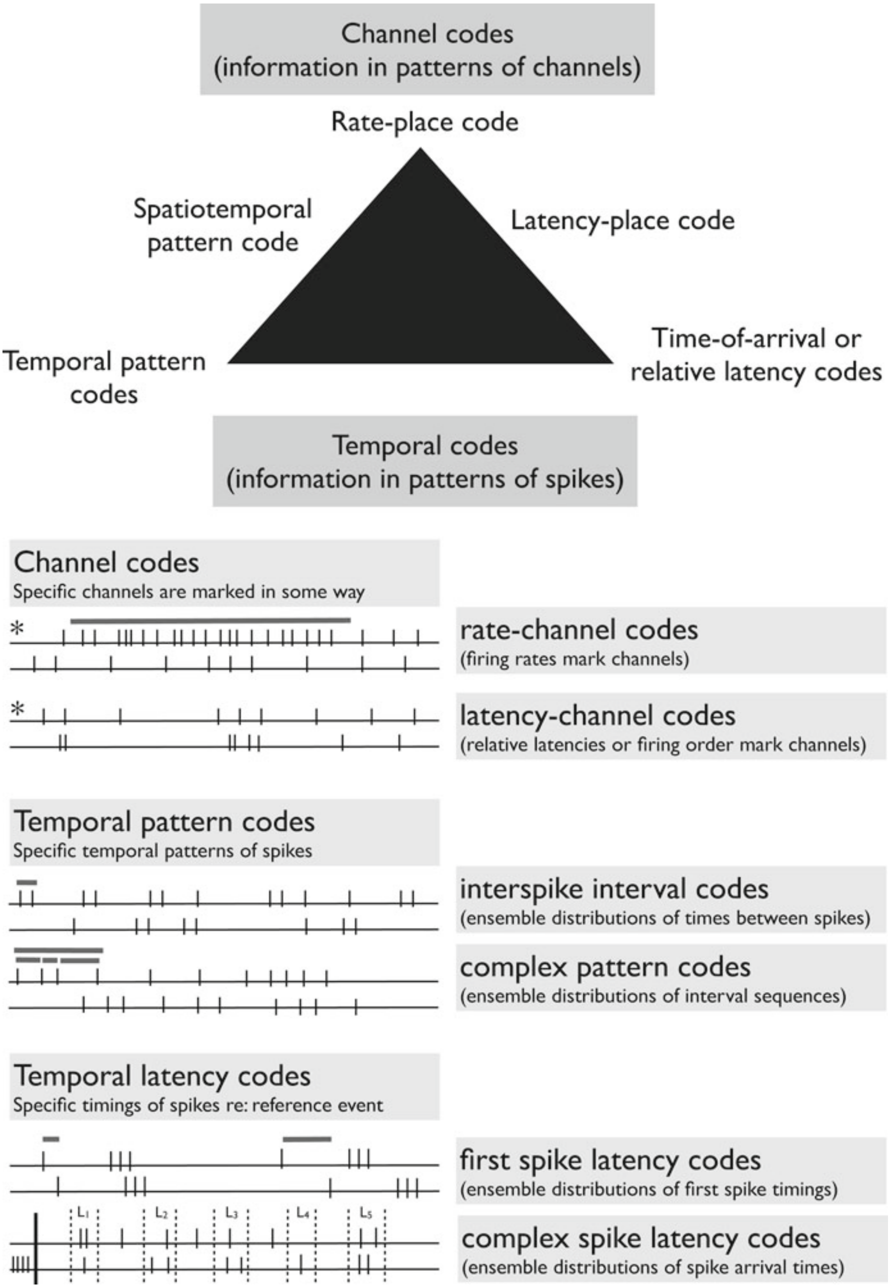

Channel codes
(information in patterns of channels)
Rate-place code
Spatiotemporal pattern code
Latency-place code
Temporal pattern codes
Time-of-arrival or relative latency codes
Temporal codes
(information in patterns of spikes)
Channel codes
Specific channels are marked in some way
*
*
rate-channel codes
(firing rates mark channels)
latency-channel codes
(relative latencies or firing order mark channels)
Temporal pattern codes
Specific temporal patterns of spikes
interspike interval codes
(ensemble distributions of times between spikes)
complex pattern codes
(ensemble distributions of interval sequences)
Temporal latency codes
Specific timings of spikes re: reference event
first spike latency codes
(ensemble distributions of first spike timings)
L1
L2
L3
L4
L5
complex spike latency codes
(ensemble distributions of spike arrival times)

Fig. 13.4 Basic types of neural pulse codes. (Top) Division of codes into channel codes, temporal pattern codes, and relative latency codes. The three types are complementary, such that combination codes can be envisioned (e.g., marking of channels by particular spike patterns, spike latencies, or firing order rather than average rate). (Bottom) Schematic illustration of different code types. Channel codes convey information via patterns of marked channels (*), whereas temporal codes

far the dominant neural coding assumption throughout the history of neurophysiology (Boring, 1942), and, consequently, forms the basis for most of our existing neural-network models.

Within channel-coding schemes, aspects of the neural response other than rate, such as relative latency or temporal pattern, can also play the role of encoding alternative signal states Combination latency-place and spatiotemporal codes are shown in Fig. 13.4. In a simple latency-channel code, channels producing spikes at shorter latencies relative to the onset of a stimulus indicate stronger activation of tuned elements. Patterns of relative first-spike response latencies can encode stimulus intensity, location, or other qualities (Eggermont, 1990; Brugge et al., 1996; Heil, 1997). Temporal, channel-sequence codes have also been proposed in which the temporal order of neural response channels conveys information about a stimulus (Van Rullen & Thorpe, 2001).

Common-response latency, in the form of interchannel synchrony, has been proposed as a strategy for grouping channels to form discrete, separate objects (Singer, 2003). In this scheme, rate patterns across simultaneously activated channels encode object qualities, whereas interchannel synchronies (joint properties of response latencies) create perceptual organization, which channels combine to encode which objects. The concurrent use of multiple coding vehicles, channel, rate, and common time-of-arrival permits time-division multiplexing of multiple objects. Still, other kinds of asynchronous multiplexing schemes are possible if other coding variables, such as complex temporal patterns and temporal pattern coherences, are used (Emmers, 1981; Cariani, 2004; Panzeri et al., 2009).

13.3.1.2 Temporal Coding Schemes

Characteristic temporal discharge patterns can also convey information about stimulus qualities. Neural codes that rely predominantly on the timings of neural discharges have been found in a variety of sensory systems (reviews: Cariani, 1995, 1999, 2001b, 2004). Conceptually, these temporal codes can be divided into time-of-arrival and temporal-pattern codes (Fig. 13.4).

Fig. 13.4 (continued) convey information via patterns of spikes (bars). In rate-place codes, across-neuron firing rate patterns of tuned elements convey information (e.g., coding of stimulus power spectra via rate profiles of frequency-tuned neurons). Temporal pattern codes use temporal patterns among spikes, such as distributions or sequences of interspike intervals, to convey information (e.g., coding of periodicity pitch via all-order interspike interval distributions). Time-of-arrival or relative latency codes use relative arrival times of spikes to convey information (e.g., coding of azimuth via spike timing disparities between left and right auditory pathways). They also can use distributions of timings of spikes with different latencies following some common initial event (vertical bar) to encode the selective activations of various neuronal assemblies that have different response latencies (L_1–L_4)

Time-of-arrival codes use the relative times of arrival of spikes in different channels to convey information about the stimulus. Examples of time-of-arrival codes are found in many sensory systems that utilize the differential times of arrival of stimuli at different receptor surfaces to infer the location of external objects (von Békésy, 1967; Carr, 1993). Strong examples are auditory localizations that rely on the time-of-arrival differences of acoustic signals at the two ears, echolocation range findings that rely on time-of-arrival differences between emitted calls and their echoes, and electroceptive localizations that use the phase differences of internally generated weak electric fields at different locations of the body to infer the presence of external phase distortions caused by nearby objects.

Temporal pattern codes, such as interspike interval codes, use temporal patterns between spikes to convey sensory information. In a temporal pattern code, the internal patterns of spike arrivals bear stimulus-related information. The simplest temporal pattern codes are interspike interval codes, in which stimulus periodicities are represented using the times between spike arrivals. More complex temporal pattern codes use higher-order time patterns consisting of interval sequences (Emmers, 1981; Abeles et al., 1993; Villa, 2000). Like time-of-arrival codes, interval and interval-sequence codes could be called correlational codes because they rely on temporal correlations between individual spike-arrival events. Temporal pattern codes should be contrasted with conceptions of temporal coding that rely on temporal variations in average discharge rate or discharge probability. These temporal-rate codes count spikes across stimulus presentations as a function of time and then perform a coarse temporal analysis on changes in spike rates.

Both time-of-arrival and temporal-pattern codes for conveying sensory information depend on spike timing patterns that are characteristic of a given stimulus attribute. The stimulus-related temporal discharge patterns on which temporal-pattern codes depend can arise in two ways: through direct stimulus locking and through stimulus-triggered, intrinsic-time courses of response (i.e., characteristic impulse response forms of receptors, sensory peripheries, and central neuronal assemblies). Some temporal codes permit signal multiplexing (Cariani, 1995, 2001b, 2004; Panzeri et al., 2009), such that different types of information can be transmitted concurrently over the same axonal transmission lines.

13.3.2 Neural Representations

Neural representations are patterns of neural activity that provide systematic means of encoding a set of informational distinctions. From psychoacoustic studies and our own direct experiences as listeners, there appear to be informational structures that provide for systematic representation of parameters that are associated with the basic auditory qualities of loudness, duration, location, pitch, and spatial hearing (directionality, apparent size). The concept of the neural code emphasizes the specific aspect of neuronal activity (e.g., firing rates, spike correlation patterns, relative latencies), whereas the concept of a neural representation emphasizes the systematic nature of the distinctions being conveyed. For example, the neural representation of sound direction relies on a spike timing code based on relative latency at the level of the auditory brain stem, but the cortical representation might be based on patterns of spike rates or first spike latencies across direction-tuned neural populations (Brugge et al., 1996). The existence of coherent internal representations for different aspects of sounds is inferred from the systematic nature of perceptual judgments, and the relative independence of judgments related to these different aspects. For example, two very different sounds, in terms of spectrum and periodicity, can nevertheless be compared in terms of their loudness, apparent location, and duration. Although there can weak interactions between these dimensions, for most low-frequency hearing involving speech and music, the dimensions are remarkably independent. The different, independent representations are the presumed basis for the dimensional structure of the percept space.

Although many different kinds of neurophysiological and neuroanatomical findings suggest that particular kinds of representations are likely to be utilized by the auditory system, the only way that one can reliably test whether a given pattern of neural activity serves as the informational vehicle for some perceptual or cognitive function is to attempt to predict the percepts experienced directly from the neural data. If the putative neural codes and representations can be used to successfully predict specific patterns of perception and cognition, given additional neuroanatomical and neurocomputational constraints, then this constitutes strong evidence that the system itself is utilizing information in this particular form.

13.3.3 High-Level Information-Processing Operations

To realize perceptual functions, the auditory system carries out operations on the encoded information in representations. These operations transform patterns of neural activity that bear information (codes and representations) into decisions that then select subsequent action. Sets of information processing operations thus realize perceptual and cognitive functions such as

- detections (e.g., judgment of presence or absence of a sound pattern, “feature detections” of all sorts)
- comparisons or discriminations (e.g., estimating the similarity of two sounds, how they are alike or different, distinguishing two different pitches or note durations)
- classification and recognition (e.g., classification of phonetic elements, recognition of a familiar word or voice)
- anticipatory predictions (e.g., producing expectations of what sounds will occur next based on what has occurred before, on multiple timescales)
- object/stream formation and separation (e.g., hearing out individual voices from a mixture)
- tracking of objects/streams (e.g., following the apparent direction of a moving sound source)

- attention (e.g., focusing on particular objects, streams or aspects of sounds, enhancement of some representations and suppression of others)
- "evaluation," (e.g., assessing the relevance of a given sound to survival, reproduction, or more specific, current goals)

These information processing operations are carried out by neurocomputational mechanisms that are discussed in the next section. Many mechanisms for comparing sound patterns and attributes involve memory mechanisms on different timescales.

13.3.4 Low-Level Neurocomputations

Neurocomputations are processes on the level of individual neurons that carry out the most basic signal processing operations. Examples of basic neurocomputations include:

- thresholding operations (e.g., spike generation)
- spike addition and temporal integration (i.e., spatial and temporal summation of excitatory inputs)
- subtraction (excitatory vs. inhibitory inputs)
- coincidence detection (spike multiplication)
- anticoincidence (spike disjunction via coincidence detection with excitatory and inhibitory inputs)
- time delay (synaptic delay, conduction delay, inhibitory rebound)
- membrane threshold accommodation (high-pass filtering, onset detection)
- spike-pattern generation (e.g., bursting patterns)
- axonal spike-train filtering (via activity-dependent conduction blocks)
- synaptic functional modification (e.g., spike-timing-dependent plasticity)

Out of these and other biophysical processes, signal processing elements such as leaky integrators, coincidence detectors, and onset detectors can be constructed. From combinations of these basic sets of computational primitives, even more complex operations can be realized.

A useful, concrete example can be found in the basic neurocomputational operations in the auditory brain stem that subserve auditory localization from acoustic interaural time-of-arrival differences. Here the neural code is a spike timing (relative latency) code that is a consequence both of differences in sound arrival time at the two ears and of phase-locking of low frequency components. A binaural cross-correlation operation is carried out in the auditory brain stem using axonal delay lines, precisely timed inhibition, and coincidence detection in bilaterally symmetric, bipolar neurons. Thus, utilizing biophysical mechanisms and dedicated, specialized neuroanatomical structures, this neural architecture implements a binaural cross correlation operation that supports systematic representation of the horizontal plane in auditory space.

13.3.5 Neural Architectures

A neural architecture is an organization of neural elements, including their interconnections and element response properties, which provide the anatomical and physiological substrate needed to implement basic neurocomputations. In turn, these neurocomputations realize information-processing functions, and ultimately perceptual and cognitive functions. Each kind of neural coding scheme requires a compatible neural architecture (what circuit organizations and element properties are available) for its implementation. Thus the question of the nature of neural codes that are operant in auditory cortex is intimately related to the question of the nature of the neurocomputations that are realized by the neural populations in auditory cortex.

What kind of neural information processing architecture is auditory cortex? There are several broad alternatives: rate-place connectionist architectures with or without spatial maps, synchronized or oscillatory connectionist architectures, time delay neural networks, or timing nets. The relative uniformity of cortical organization suggests that one basic architectural type handles all different kinds of incoming information, albeit with plastic adjustments that depend on the correlation structure of the inputs. Within the constraints given by coarse genetic specifications, the stimulus organizes the fine structure of the tissue. In all of these neural network types, network functions can be adaptively modified by changing synaptic efficacies and other biophysical parameters. Given this plasticity, it is almost certain that auditory cortex configures itself in different ways according to the different kinds of information that are determined by connections to other parts of the system (sensory surfaces, subcortical afferent pathways, descending pathways, other cortical and subcortical populations). It is conceivable that auditory cortex can support several, perhaps all, of these alternative processing organizations.

13.3.5.1 Connectionist Architectures

Rate-place connectionist architectures are neural networks in which all processing involves analysis of firing rate profiles among neural channels (“units” or “nodes”). The cerebral cortex is commonly regarded as a large recurrent connectionist Hopfield network whose informational states are N-dimensional vectors that represent the firing rates of its neural elements (e.g., Trappenberg, 2002). Because firing rates are scalar quantities, all informational distinctions must be made via different combinations of neural channel activations. Thus, the most basic assumption of connectionist systems is channel-coding (which channels are activated how much).

In a connectionist network, the signals emitted by each channel are “labeled” by virtue of their specific intranetwork connectivities that in turn determine their simple and complex tuning properties. In auditory cortex, neurons are thought to convey different kinds of information depending on their various tuning properties, such as selectivity for

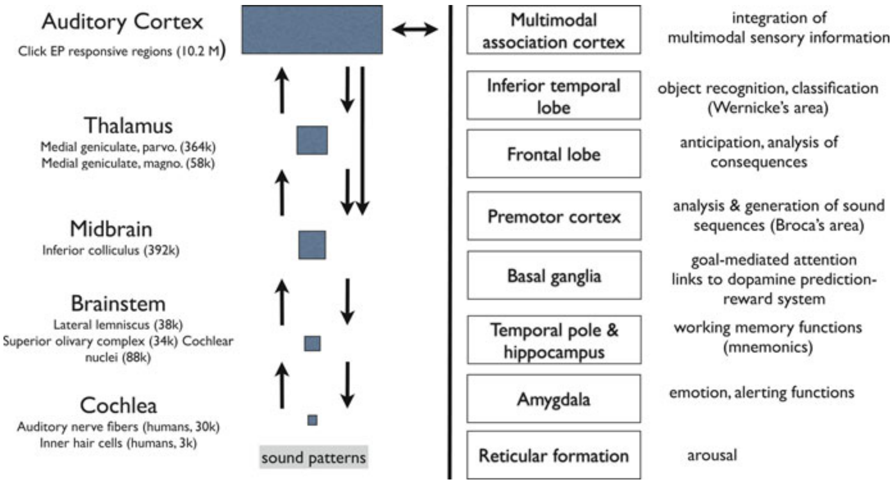

Auditory Cortex
Click EP responsive regions (10.2 M)
Thalamus
Medial geniculate, parvo. (364k)
Medial geniculate, magno. (58k)
Midbrain
Inferior colliculus (392k)
Brainstem
Lateral lemniscus (38k)
Superior olivary complex (34k) Cochlear nuclei (88k)
Cochlea
Auditory nerve fibers (humans, 30k)
Inner hair cells (humans, 3k)
sound patterns
Multimodal association cortex
integration of multimodal sensory information
Inferior temporal lobe
object recognition, classification (Wernicke's area)
Frontal lobe
anticipation, analysis of consequences
Premotor cortex
analysis & generation of sound sequences (Broca's area)
Basal ganglia
goal-mediated attention links to dopamine prediction-reward system
Temporal pole & hippocampus
working memory functions (mnemonics)
Amygdala
emotion, alerting functions
Reticular formation
arousal

Fig. 13.5 Functional connections of auditory cortex to auditory pathway and the rest of the brain. (Left) Major levels in ascending and descending auditory pathways. Except where noted, numbers associated with auditory structures indicate numbers of neurons in the squirrel monkey auditory system (Chow, 1951). (Right) Major projections between auditory cortex and other brain structures, along with their basic functions. Connections between these structures and subcortical auditory pathways have been omitted

frequency, periodicity, sound location, bandwidth, intensity, or their modulations. The various tuning properties of a given neuron in turn depend on how it is connected to other neurons in the auditory system and other parts of the brain (Fig. 13.5).

In this vein, a number of studies have used linear system-identification techniques to characterize the time-frequency tuning properties of cortical neurons in terms of spectrotemporal receptive fields (STRFs) (e.g., Miller et al., 2002). Ideally, one should be able to use STRFs to predict the running firing rates of characterized neurons, ensembles, and populations to novel, complex stimuli. In practice, many neural elements behave unpredictably, with nonlinear responses that can change dynamically depending on recent stimulus history (Fritz et al., 2003).

There has been an ongoing debate about the nature of cortical processing elements, whether they are rate-integrators that are more compatible with connectionist schemes, or coincidence detectors operating on some kind of temporal code. If the functioning of auditory cortex does in fact depend on channel-coding schemes that use elements with relatively fixed receptive fields, be they dense or sparse, one might a priori expect the elements to have more reliable behavior (less discharge rate variance). On the other hand, all optimality arguments about neural codes and architectures are very risky to invoke at this point, before a reasonably firm grasp of how the system works has been attained. Given multitudes of neural elements, pooling of firing rate information via statistical population codes could potentially reconcile this apparent incongruity (see Section 13.4.2), but concrete mechanisms for pooling this information have yet to be identified. So far, no definitive answer has emerged.

Perhaps even more challenging for connectionist networks are problems of simultaneously representing and analyzing multiple auditory objects and event streams. More flexible kinds of networks are clearly needed to handle the combinatorics of multiple objects and their associated attributes. Temporal correlations between spike patterns (von der Malsberg, 1994) and emergent synchronies between spikes (Singer, 2003) could serve to bind together various feature detector channels that would group together corresponding attributes of auditory and visual objects. Synchrony-based grouping mechanisms have been the focus of much neurophysiological study in the visual cortex, albeit with equivocal correspondences with perception. Along similar lines, synchronized oscillations of neural firing have been proposed as auditory mechanisms for grouping channel-coded features and separating multiple sounds (Wang, 2002).

13.3.5.2 Oscillatory Networks

Neuronal oscillations have long been considered as potential mechanisms for informational integration (McCulloch, 1951; Greene, 1962). Stimulus-driven, stimulus-triggered, and endogenous, intrinsic oscillations are widespread in the brain (Buzsáki, 2006). Stimulus-driven oscillations follow the time structure of the stimulus, whereas stimulus-triggered oscillations, although evoked by an external stimulus, have their own intrinsic time courses that can also convey information about the stimulus (Bullock, 1992; Thatcher & John, 1977). Emergent, stimulus-triggered oscillations have been observed in olfactory systems and in the hippocampus, where spike latencies relative to oscillatory field potentials plausibly encode respectively, odor qualities (Laurent, 2006) and positional information relevant for navigation. These kinds of phase- or latency-based codes can either support marking of specific subsets of channels or ensemble-wide readouts of complex temporal patterns of response latency (Fig. 13.4). General purpose oscillatory-phase-latency codes for encoding signals and rhythmic-mode processing mechanisms for integrating multimodal information have been proposed (Schroeder & Lakatos, 2009).

Despite widespread evidence for oscillatory coupling of many neuronal populations it is not yet clear whether the various gamma, theta, and alpha oscillations that are seen in cortical populations play obligatory or specific informational roles as either temporal frameworks for phase-precession codes or channel-grouping mechanisms. Instead, the oscillations might be general signs of neuronal activation that co-occur when neurons are excited and information is being processed, but have little or no specific informational function. For example, gamma rhythms in cortical populations are reflections of excitatory and inhibitory dynamics of pyramidal and basket cells that appear when cortical pyramidal cells are maximally driven, but there appears to be little or no information conveyed in specific oscillatory frequencies. In some cases stimulus detection thresholds are lower when stimulus presentations are timed to coincide with recovery phases of oscillations, but this may simply reflect the larger numbers of neurons available and ready to respond at those moments. Here oscillations play a somewhat more tangential, facilitating role vis-à-vis neural coding and information processing.

Perhaps the field can learn from its history. In the past an intriguing "alpha scanning" mechanism was proposed as a substrate for computing form invariants (McCulloch, 1951), but this hypothesis was severely undermined by the relative ease that alpha rhythms can be disrupted at will without major perceptual or cognitive consequences. Today critical experiments likewise need to determine whether phase-resets or abolition of oscillations using appropriately timed stimuli, such as clicks, flashes, shocks, or pharmaceutical interventions can significantly disrupt functions. Experiments along these lines could clarify what dependencies exist between neural information processing mechanisms and the stimulus-driven, stimulus-triggered, and intrinsic oscillatory neurodynamics of neuronal excitation, inhibition, and recovery.

13.3.5.3 Time-Delay Neural Networks, Synfire Chains, and Timing Nets

Thus far, both traditional connectionist networks and synchronized, oscillatory, and/or temporally gated connectionist network assume channel coding of specific stimulus attributes. In the early auditory system, however, many stimulus distinctions appear to be conveyed by means of temporal codes.

Time-delay neural networks can be used to interconvert time and place (channel) patterns. In essence, any fixed spatiotemporal spike volley pattern can be recognized and produced by implementing appropriate offsetting time delays within and/or between neural elements. Classical time-delay networks used systematic sets of synaptic and axonal transmission delays embedded in arrays of coincidence detectors to convert temporal patterns to activations of specific channels. These include temporal correlation models for binaural localization (Jeffress, 1948), periodicity pitch (Licklider, 1959), and binaural auditory scene analysis (Cherry, 1961).

Modulation-tuned elements can be also used to convert time to place, and periodotopic maps consisting of such elements have been found in the auditory pathway (Schreiner & Langner, 1988). These maps form modulation spectrum representations of periodicities below 50 Hz that can usefully subserve recognition of consonantal speech distinctions and rhythmic patterns. Although neural modulation spectra have been proposed as substrates for periodicity pitch, modulation-based representations for pitch break down when confronted with concurrent harmonic tones (e.g., two musical notes a third apart).

Synfire chains (Abeles, 2003) and polychronous networks (Izhikevich, 2006) are time-delay networks in which spatiotemporal channel activation sequences are propagated. These are distinct from both connectionist and time-delay networks in that both channel and timing are equally important. Information is encoded in the spatiotemporal trajectory of spikes through the system. Because each trajectory depends on specific interneural delays and synaptic weightings, it is unclear how stimulus invariances and equivalences might be realized this way. However, one of the major potential advantages of synfire and polychronous networks is their ability to multiplex signals. In these networks a given neuron can participate in multiple synfire chains and polychronous patterns, and this mutual transparency of signals drastically simplifies the neurocomputational problems of representing multiple attributes and objects.

Timing nets are a third general type of neural network that are distinct from both connectionist networks and time-delay networks (Cariani, 2001a, 2004). Whereas connectionist networks operate entirely on channel activation patterns, and time-delay networks convert temporal patterns into channel activations, timing nets operate entirely in the time domain. Timing nets are similar to time-delay neural networks in that they consist of arrays of coincidence detectors interconnected by means of time delays and synaptic weights. Whereas both types of networks have temporally coded inputs, the outputs of timing nets are also temporally coded rather than by channel.

Simple timing nets have been proposed for analysis of periodicity and spectrum and for grouping and separation of auditory objects. Feedforward timing nets act as temporal pattern sieves to extract common spike patterns among their inputs, even if these patterns are interleaved with other patterns. Such operations elegantly extract common periodicities and low-frequency spectra from two signals, for example, recognizing the same vowel spoken by two speakers with different voice pitches (different fundamental frequencies [F0s], same spectra) or different vowels spoken by the same speaker (different spectra, same fundamental frequencies). Such networks can also be used to separate out and recognize embedded and interleaved temporal patterns of spikes, an important property for multiplexing of multiple temporal pattern signals and for complex, multidimensional temporal representations. Timing nets illustrate how processing of information might be achieved through mass statistics of spike correlations rather than through highly specific connectivities.

Recurrent timing nets consist of delay loops and coincidence elements that carry circulating temporal patterns associated with a stimulus (Cariani, 2001a, 2004; see also the recurrent neural loop model of Thatcher & John, 1977). The nets in effect multiply a signal by its delayed version to build up and separate multiple repeating temporal patterns that are embedded in the signal. The auditory system readily separates multiple musical notes whose fundamental frequencies (F0s) are separated by more than 10% (e.g., nonadjacent notes on the piano). Such note combinations have embedded within their waveforms two different patterns that have different repetition times (fundamental periods). The time-domain filtering operations carried out by the delay loops act roughly like comb filters to produce two sets of signals that resemble the individual vowel waveforms. In neural terms, they separate the two vowels on the basis of invariant temporal patterns of spikes rather than by segregating and binding subsets of activated periodicity or spectral feature channels. In doing so, they provide an example of how auditory object formation based on harmonic, periodic structure could occur at very early stages of auditory processing, before any explicit frequency and periodicity analysis takes place. On larger timescales, such networks can build up and separate repeating, complex rhythmic patterns as well (Cariani, 2002).

Feedforward and recurrent timing nets were developed with temporal coding of pitch and auditory scene analysis in mind. Because they operate on temporal patterns of spikes that are not evident at the level of auditory cortex, neural timing net mechanisms for periodicity pitch analysis and F0-based sound separation would likely need to be located at earlier stages of auditory processing, possibly dynamically facilitated by descending projections to thalamus and midbrain (see discussion of reverse-hierarchy theory in Section 13.3.6.2). Because coarser temporal patterns of spikes associated with onsets and offsets of auditory events are present in cortical stations, recurrent timing mechanisms could exist at those levels to carry out coarser temporal pattern comparisons whose violations produce mismatch negativities.

13.3.6 Functional Roles of the Auditory Cortex

In considering the functional role of the human auditory cortex vis-a-vis the rest of the brain, it is useful first to summarize some general principles that govern brain organization and function.

13.3.6.1 General Principles of Brain Organization and Function

In cybernetic terms, brains can be seen as adaptive, goal-directed percept-action systems. Sensory systems gather information about the surrounding world (sensory functions). Cognitive representations and operations evaluate incoming sensory inputs and prospective actions in the context of previously acquired knowledge. Motor systems carry out actions on the world (motor functions). Coordinative linkages, from simple reflex arcs to much more complex circuits, link percepts and cognitive representations to actions. Motivational goal systems steer perception and action toward satisfaction of immediate needs, while anticipatory and deliberative systems analyze the deeper ramifications of sensed situations and plan prospective actions (executive functions) that satisfy longer range goals. Evaluative reward systems judge the effectiveness of sensorimotor linkages vis-à-vis goals and adaptively modify neural subsystems to favor behaviors that fulfill drive goals to avoid those that are detrimental to survival. Affective and interoceptive systems provide a running estimate of the state of the organism that influence choice of behavioral alternatives (e.g., fight/flight). Mnemonic systems retain associations between sensory information, internal deliberations, sensorimotor sequences, and rewards for later use by steering mechanisms that take into account anticipated consequences of action alternatives (rewards and punishments).

These different functionalities are subserved by different subcortical and cortical neuronal populations (Mesulam, 2000). Cerebral cortical regions are involved in sensory, motor, coordinative sequencing, anticipatory, and executive functions. The cerebellum involves real-time motor adjustments and control of sensory surfaces. Hypothalamus and amygdala are involved with fixed drives and affect-based modulation of behavior. Dopaminergic predictive reward circuits reconfigure the system to incorporate new goals. Basal ganglia structures steer attention and switch action modes to address current, salient goals, providing linkages between limbic-generated goal states and cortical sensorimotor processing.

Some basic principles exist for cortical organization. Within general neuroanatomical plans that are specified through genetic guidance of developmental processes, most large-scale patterns of cortical functional connectivity can be understood through the interaction of correlated external inputs, internal reward signals, existing interneural connectivities, and the action of activity-dependent biophysical mechanisms that alter them.

The first maxim is “cortex is cortex,” meaning that different cortical regions have roughly the same cell types and general organization, albeit with varying relative cell densities and connectivities among and within the cortical layers. A second is that the “stimulus organizes the tissue” such that the dominant inputs to a given region alter the fine structure and function of the tissue according to the correlational structure of its inputs and outputs vis-à-vis effective action. The functional organization of unimodal cortex is largely determined by the afferent inputs and ultimately by the organization of sensory and motor surfaces. Thus, auditory cortex has several fields that are coarsely cochleotopically organized, in parallel with retinopic organization in visual cortex, and somatotopic organization in somatosensory cortex.

A third organizing principle is that there is an ongoing competition for cortical territory that is mediated by the strength of both incoming information and internal evaluative reward signals. The strongest, most internally rewarded inputs come to dominate the responses of a given region over time. When normal sensory inputs to a patch of cortex are silenced, other weaker inputs are strengthened (by sprouting and synaptic proliferation, stabilization, and strengthening). Provided that they play a useful functional role such that they are internally rewarded, such weak inputs can then come to dominate responses.

A fourth rule-of-thumb is that connectivities between neural populations are almost invariably reciprocal, such that recurrent loops are norm rather than exception. “Everything is connected” by such recurrent loops, that is, there are multisynaptic pathways that provide reciprocal connections between any two neurons in the system. Because “neurons that fire together wire together” even arbitrary long-range reciprocal connections can be made and stabilized. Lastly, lateral interconnections are mostly local and short range. These connectivity patterns lead to cortical convergence zones that handle confluences of different types of sensory information (Damasio & Damasio, 1994), provided that the different types of information correlate in a functionally meaningful way (internally rewarded). Cortical regions that operate on similar kinds of information and/or perform similar tasks therefore tend to be clustered together spatially. Much of the large scale functional topography of cortical regions may ensue from these basic principles (e.g., dorsal paths for localization leading to body and extrapersonal space maps in the parietal lobe, ventral paths for object recognition leading to regions in the temporal lobe, hemispheric colocalizations of related, time-critical functions).

13.3.6.2 Conceptions of Auditory Cortical Function

The auditory cortex receives incoming sensory information from the ears via ascending afferent auditory pathways, and controls the information it receives through descending, efferent pathways that modulate neural activity at every level of processing (Fig. 13.5; Winer, 1992; Clarke and Morosan, Chapter 2). The auditory cortex has reciprocal connections with other cortical regions involved with object recognition and classification (temporal lobe), analysis and production of sensorimotor sequences (premotor frontal regions), expectancy and decision making (frontal regions), body space (parietal regions), as well as with uni- and multimodal cortical regions associated with other sensory systems.

By virtue of its connections to the auditory pathway and to other functionally related cortical and subcortical (limbic, basal ganglia) areas, auditory cortex is strategically situated to coordinate processing of auditory information for a number of organism-level purposes. These purposes include monitoring changes in the environment (alerting functions), separating sound objects and streams (perceptual organization), detecting and discriminating relevant sounds (discriminatory functions), recognizing familiar sounds (classificatory and mnemonic functions), locating relevant sound sources (orienting functions), decoding speech communication signals (phonetic, syllabic, and word classification and sequence analysis functions), providing feedback for sound production processes, and self-regulation of internal state (e.g., use of music to regulate mood, affect, pleasure, arousal).

Currently two broad conceptions exist concerning the role of auditory cortex vis-à-vis lower stations (perspectives often heavily shaped by whether one has investigated the system at subcortical or cortical levels). The first conceives of auditory cortex as the culmination of the auditory pathway, the stage at which all incoming auditory information is organized and analyzed. Here auditory cortex is the nexus for fine-grained representations of sound that are used for auditory functions. In this sequential-hierarchical feedforward view, it has been assumed that "higher level functions" such as recognition of phonetic tokens and the organization of the auditory scene take place at the cortical level after a basic frequency and spatial hearing analysis has been first carried out by lower stations.

A second, emerging perspective conceives of auditory cortex as a control system. In vision this has been termed "the reverse hierarchy theory" (Ahissar & Hochstein, 2004). The main purpose of such a control system is not as a repository of fine-grained representations. Rather, it is to organize information processing in "lower" circuits at thalamic, midbrain, brain stem, and even perhaps cochlear levels by means of descending connections that can release inhibitory controls. This disinhibitory control may be similar in function to the double-inhibitory mechanism by which in basal ganglia activity release inhibition to bias activity patterns in cortical motor areas toward particular actions (movement initiation, switching) and to bias sensory areas to facilitate particular signals (attention). The system in effect chooses its own inputs contingent on its immediate interests.

The representations needed for such a control system do not necessarily need to be as precise as perceptual acuities if the cortex can access fine-grained temporal information at lower stations when needed. When presented with a task requiring attention and fine discrimination, the cortex could potentially pose the question to lower levels by setting up (by disinhibition) dynamic neural linkages that facilitate and hold the informational distinctions that are needed. This theory has the merits that it is consistent with the massive descending pathways that are present in both the auditory and visual systems, and it also provides some explanation as to how fine-grained temporally coded information might be used by central stations in the auditory system, yet not be present in precise and overt form. It is consistent with relatively recent evidence that cortical activity may modulate lower level processing even as far down as the brain stem, on both short- and long-term timescales (Tzounopoulos et al., 2004; Lee et al., 2009).

13.4 Fundamental Issues and Open Problems

13.4.1 Identifying Neural Codes and Representations at the Cortical Level

Perhaps the most fundamental open problem at the cortical level is to identify the specific neural codes that subserve different perceptual and cognitive representations, such as pitch, timbre, location, and loudness (Phillips et al., 1994; Brugge et al., 1996; Furukawa & Middlebrooks, 2004; Bendor & Wang, 2005; Bizley & Walker, 2010; Hall and Barker, Chapter 7). Cortical representations related to pitch and rhythmic pattern are most important for music (Zatorre and Zarate, Chapter 10), whereas those related to timbral, phonetic distinctions are most important for speech communication (Huetz et al., 2011; Giraud and Poeppel, Chapter 9). The nature of cortical codes places strong constraints on neural mechanisms for higher-level informational integration, in the specific processes that form auditory objects and streams (Shamma & Micheyl, 2010; Griffiths, Micheyl, and Overath, Chapter 8), in the integration of auditory representations with those of other senses (van Wassenhove and Schroeder, Chapter 11), and in the utilization of auditory information for action (Hickok and Saberi, Chapter 12). This section lists and briefly describes some of the most important unresolved issues concerning the nature of auditory codes and representations that apply generally to all of the aforementioned problem domains of basic auditory constituents, music, speech, auditory scene analysis, multimodal representations, and sensorimotor integration.

13.4.1.1 Rate, Channel, and Time Codes

Because of their prominent and abundantly documented tonotopic organization, the peripheral and central auditory systems have often been conceived as an ensemble of labeled-line frequency channels, such that profiles of average firing rates across tonotopic axes provide a central, general-purpose representation of the stimulus power spectrum. Similarly, cortical units whose average firing rates covary with many other acoustic parameters, such as periodicity, intensity, duration, amplitude and frequency modulation, bandwidth, harmonicity, and location have been found. This leads to the hypothesis that representation of the auditory scene at the cortical level is simply a matter of analyzing average firing rate profiles among a relatively small number of neural subpopulations that encode feature maps. Such coding schemes work best with elements that have stable receptive fields, with sensitivity to only one or two acoustic parameters. Complicating this picture, however, is the problem of disentangling the multiple parameters that can influence any given neuron's firing rate, especially if multiple auditory objects are simultaneously present.

A strong case can be made that the central representations for both periodicity pitch and spectral determinants of timbre are ultimately based on population-wide interspike interval statistics at early stages of auditory processing (Palmer, 1992; Cariani & Delgutte, 1996; Cariani, 1999; Ando & Cariani, 2009). Although individual neurons and neuronal ensembles in lightly and unanesthetized auditory cortex can phase lock up to stimulus periodicities of several hundred Hz (Fishman et al., 2000; Wallace et al., 2002), most cortical neurons do not go above 30–40 Hz (Miller et al., 2001). Thus, the direct, iconic temporal-pattern codes for pitch and timbre that predominate in the auditory periphery and brain stem appear to be largely absent at the cortical level, necessitating some form of coding transformation (Wang, 2007). The most specific neural correlates of pitch found to date in auditory cortex instead involve specialized subpopulations of neurons whose firing rates are tuned to particular periodicities (Bendor & Wang, 2005). Questions of how peripheral timing patterns might be transformed in the central auditory system to give rise to such cortical pitch detectors are still unresolved.

However, other types of temporal codes that are based on the relative latencies of spikes rather than stimulus-driven temporal patterns are possible at the cortical level. Neurons in A1 appear to encode stimulus onset timing very precisely in their response latencies (Heil, 1997; Phillips et al., 2002). Representations can be based on latency differences across units (i.e., latency-place) codes, or dynamic latency-coding schemes (Heil, 1997). For example, the loudness of an abrupt, short duration tone can be encoded by the temporal dispersion of first-spike responses over a population. Multiplexed sparse distributed temporal codes (Abeles et al., 1993; Villa, 2000; Panzeri et al., 2009) in which periodicity-related spikes are interspersed with those encoding other kinds of perceptual information (timbre, spatial attributes) may exist in auditory cortex (Chase & Young, 2006) in some covert form that is difficult to recognize. Some evidence exists for precise temporal sequences of spikes that are related to perceptual functions (Villa, 2000). Because the latency of these sequences can vary from trial to trial, they may be smeared in poststimulus time histograms.

13.4.1.2 Sparse-Efficient versus Abundant-Redundant Codes

With the advent of information theory and its application to neuroscience and psychology, the degree of redundancy of neural responses at various levels of processing within sensory system has become a key issue in the analysis of neural representations. Horace Barlow proposed that neural representations of stimuli become less and less redundant at each successive processing stage within sensory systems (Barlow, 1961). In this context, the question of whether representations at the level of auditory cortex are in some sense less redundant than the representations at lower levels of the auditory system has been raised (Chechik et al., 2006).

Another important question related to coding redundancy concerns the "sparseness" of neural representations. One way to characterize sparseness involves counting how many neurons in a population are active during the presentation of a stimulus, and how many are quiescent (Hromadka et al., 2008, Bizely et al., 2010). If only a relatively small number of neurons are active (e.g., <10%), the neural representation of the considered stimulus is said to be sparse. Another approach to sparseness involves counting how many spikes each neuron produces in response to a stimulus. In theory, sparse representations are desirable because they are more energetically efficient. The downside, of course, is reduced resilience to individual-component failure, or malfunction.

If sound representations in auditory cortex are efficient, and sparse, one may wonder why there should be so many more neurons at the cortical level, compared to lower stations in the auditory system. One possible answer to this question is that auditory cortex has many other functions besides the efficient representation of sound. In particular, it may have to perform complicated computations on multiple auditory representations that in turn need to be registered and coordinated with information provided by other sensory modalities (DeWeese et al. 2005, van Wassenhove & Schroeder, Chapter 11). Several recent studies have identified neurons in auditory cortex whose responses are modulated by nonauditory influences (Bizley & King, 2008; Kayser et al., 2008; Panzeri et al., 2009).

While questions of how and to what extent the redundancy of neural representations vary as one ascends the auditory pathway, perhaps the more fundamental question is why this should be so in the first place. From a functional point of view, lower redundancy makes for more efficient coding in an information-theoretic sense. On the other hand, in the face of abundant sources of both internal and external noise, redundancy also plays a critically important role in enhancing reliability. Therefore, one would expect a well designed neural information–processing system to achieve a judicious balance between efficiency and redundancy.

It is possible that Barlow's coding hypothesis is not testable given our current level of understanding of neural coding at the cortical level. A pervasive problem with optimality arguments in biology is that one does not know a priori for what specific functions the system has been optimized, and what constraints (structural, developmental, evolutionary) have shaped it. Optimality analysis will rest on much firmer ground once the basic operating principles of the system (codes, computations, functions) are better understood and various design trade-offs can be more realistically assessed.

13.4.1.3 Coding of Features versus Objects

Neurons in primary and secondary auditory cortex have been found to respond in a selective manner to various sound "features," such frequency sweeps (Tian & Rauschecker, 1994), bandwidths (Rauschecker & Tian, 1994), or temporal and/or spectral modulation rates (Kowalski et al., 1996). However, many of these features are already extracted and represented in some way in lower stages of the auditory system. Thus, even though some important differences have been identified between cortical and lower-level responses (e.g., in the broadness of tuning, the nonmonotonicity of rate-level functions), it is tempting to think that there must be more to auditory cortex function than just the extraction and representation of disjoint features. This leads to the notion that auditory cortex may be a place where representations of various sound features are conjoined in a meaningful way to form representations of auditory objects (Nelken et al., 2003). Empirical evidence for the representation of auditory objects, or streams, and not just features at the level of auditory cortex, however, still remains very limited. One line of evidence comes from the results of several single-unit electroencephalography (EEG), magnetoencephalography (MEG), and functional magnetic resonance imaging (fMRI) studies, which concur to indicate that neural responses in primary and/or secondary auditory cortex reflect auditory streams (Shamma & Micheyl, 2010; Shamma et al., 2011; Griffiths, Micheyl, and Overath, Chapter 8). Another line of evidence that neural responses in auditory cortex reflect not just physical stimulus properties, but also the perceptual organization of these features into objects, comes from EEG studies that have identified a wave (the "object-related negativity"), which appears to depend specifically on whether listeners hear out a mistuned component in an otherwise harmonic complex as a separate object (Alain and Winkler, Chapter 4). Although these findings provide important hints that auditory cortex does indeed represent auditory objects, additional research is needed to clarify the neural mechanisms whereby representations of different sound features are combined to form representations of auditory objects at the level of auditory cortex.

13.4.2 The Hyperacuity Problem

For many perceptual discriminations, the most highly tuned receptive fields of neural elements are typically much coarser—by one to two orders of magnitude than the finest distinctions that can be made by the organism as a whole. The problem of accounting for this apparent discrepancy, which exists in nearly every sensory modality, is known as the hyperacuity problem (Rieke et al., 1997). A striking example of hyperacuity problem in the auditory modality relates to the relationship between neural frequency selectivity and behavioral frequency discrimination. Just-noticeable differences (JNDs) in the frequency of moderate-level pure tones below 2 kHz can be as small as 0.1–0.2% (Moore, 1973). At 1 kHz, this corresponds to a frequency difference of about 1 Hz. In contrast, at moderate sound levels the firing-rate response bandwidths of auditory neurons at all stations in the pathway are typically on the order of large fractions of an octave (Evans et al., 1992). Although there have been recent reports of "ultrafine" frequency tuning at the single-unit level in human auditory cortex (Bitterman et al., 2008), the frequency JNDs that were estimated based on such tuning (around 3%) are still an order of magnitude larger than the smallest JNDs that can be achieved by human listeners. The usual "solution" to this discrepancy assumes that the latter do not rely on rate-place representations, but rather on temporal information, that is, phase locking. However, because phase-locking decreases sharply at successively higher auditory stations (Cariani, 1999), this type of explanation is unlikely to apply at the level of auditory cortex. Thus, either behavioral frequency discrimination is determined below cortex, or sufficiently precise neural representations of pure-tone frequency must exist at the level of auditory cortex that can account for the exquisitely small JNDs that are observed in humans and other animals.

13.4.3 The Invariance Problem

Typically, many auditory attributes can be highly invariant with respect to changes in sound parameters. A prime example is perceptual invariance of low-frequency sounds with respect to stimulus intensity. Although the loudness of sounds invariably increases monotonically as a function of sound pressure level, for low-frequency sounds the same sound presented at different levels is recognizably similar in pitch, timbre, duration, and location. For high-frequency tones, however, pitch and timbre are much more labile.

These perceptual invariances are obtained despite profound changes in both absolute and relative neural firing rates at all levels of auditory processing. Cortical sound-responsive neurons with nonmonotonic rate-level functions are quite common, which greatly complicates population-based explanations of level-invariant percepts and equivalence classes (Tramo et al., 2005). It is one of the main reasons that coherent rate-based tonotopic spatial organization is only seen only at low sound pressure levels near neural response thresholds and breaks down at higher levels (Phillips et al., 1994). Ironically, level-invariant, rate-based frequency tunings have been observed in marmoset cortex for high-frequency pure tones (Sadagopan & Wang, 2008), the very stimuli for which human pitch percepts are the least invariant with respect to level.

A second example of invariance is the relative stability of pitch, timbre, loudness, and location with respect to sound duration. This stability generally holds for durations longer than 50–100 ms. For shorter time periods, pitch strength, timbre, and loudness can change dramatically with duration. A third major invariance is the relative stability of pitch, timbre, loudness, and duration with respect to sound-source location relative to the listener.

Related to perceptual invariances are perceptual equivalence classes. Sounds consisting of low-frequency, resolved harmonics that have different phase spectra (and consequently waveform envelopes) nevertheless are indistinguishable. Harmonic, low-frequency sounds having the same fundamental frequency almost invariably produce the same low pitch at the fundamental, despite profound differences in spectral content. Pitch equivalence classes are especially important in music, where various instruments with differing spectral and dynamic characteristics play the same notes that evoke the same pitches. This pitch equivalence is what permits different types of instruments to readily serve as tuning references for each other. Octave equivalences produce pitch chromas that form the foundation for tonal pitch classes in music theory. That these same broad pitch equivalence classes extend to fundamental frequencies well beyond the range of human voices, and that they are shared by a phyletically broad range of animal listeners strongly suggests that they are integral products of basic auditory mechanisms for analysis and separation of sounds rather than the products of ontogenetic associative learning or recent evolution. At the level of the auditory nerve, pitch and octave equivalence falls out of common features in all-order interspike interval codes (Cariani & Delgutte, 1996; Cariani, 1998, 2002), whereas at the cortical level pitch equivalence may be manifested by the responses of periodicity-tuned neurons (Bendor & Wang, 2005).

13.4.4 The Transformation Problem

In vision, within limits, shapes remain perceptually invariant, such that they can be recognized when translated, rotated, and magnified with respect to retinal coordinates. This was known to the Gestaltists as form invariance under transformation. Despite the large changes that occur in retinotopic patterns of activity, the representations of these shapes nevertheless retain essential, relational aspects that are used to judge similarity and to support recognition. In audition and the temporal sense, three analogous invariances exist for pitch relations, timbral relations, and temporal event relations. These are, respectively, transpositional invariance for melodies and chords, timbral invariance of for vowels spoken by different speakers, and tempo invariance for rhythmic patterns.

Melodies are temporal sequences of pitched-events. Transpositional invariance is illustrated by the common observation that musical melodies can be identified even after they are transposed into a different key or register (frequency range). Transpositional invariance involves the ability to recognize a melody on the basis of relative pitch relations, irrespective of the absolute fundamental frequency of the beginning note. The operation of transposition multiplies all frequencies by a constant factor, thereby retaining the same frequency ratios and proportionalities. Recognition of transposed melodies is highly reliable if the melody is familiar and/or harmonically well structured (i.e., “tonal”), and transposed notes all bear the same frequency ratios (i.e., in musical terms, if musical intervals are preserved), but is much weaker and conditional if only pitch contours (patterns of up–down transitions of successive pitches) are retained (Handel, 1989; McDermott & Oxenham, 2008).

Chords can also be transposed. Chords are multiple notes played together. The type of a chord (e.g., major vs. minor vs. diminished or augmented) is determined by the musical intervals (frequency ratios) between its constituent notes. With a little exposure, human listeners can distinguish different types of consonant and dissonant chords irrespective of the absolute note frequencies that constitute them. The existence region for transpositional invariance of melodies and chords parallels that for musical tonality. Transpositional invariance, being based on musical intervals, appears to be associated with periodicity pitch, and may therefore ultimately depend on properties of temporal, interspike interval codes for periodicity pitch in early auditory processing.

Timbral invariance involves ability to recognize common timbral qualities despite changes in absolute acoustical parameters. Perception of phonetic distinctions in speech is relatively invariant with respect to the considerable acoustical variations that are produced by different speakers with different vocal tract sizes. In early studies of vowels, phoneticists found that male and female productions of the same, perceptually equivalent vowels have different absolute formant frequencies, but relatively more similar formant ratios. Interestingly, sensitivity to formant ratios has recently been observed in MEG responses to synthetic vowels in auditory cortex (Monahan & Idsardi, 2010). Vowel normalization is an operation that produces a more invariant representation by taking into account formant ratios (F2/F1, F3/F2, F3/F1) and/or formant-voice pitch ratios (F1/F0, F2/F0, F3/F0). In the auditory nerve, the most intense harmonic in each formant region dominates the interspike intervals that are produced ("synchrony capture"), such that the temporal representation of vowels resembles that produced by a small number of harmonically related pure tones (Delgutte & Kiang, 1984). The formant frequency ratios that may determine the different timbral categories of vowels are thus not unlike the tonal frequency ratios that constitute different musical intervals and chords (see also the timbral intervals discussed in McAdams & Giordano, 2009). Thus, similar kinds of mechanisms conceivably subserve the transpositional invariances of musical intervals, chords, melodies, and even vowel timbres.

Tempo invariance involves the ability to recognize a rhythmic or melodic pattern when played at different speeds. As long as the time intervals between notes are neither too short nor too long (roughly, $0.1 \text{ s} < I < 2 \text{ s}$), the temporal pattern invariance holds as long as the time intervals are all changed proportionately.

Invariance under transformation is a fundamental unsolved problem for computational neuroscience (von der Malsberg, 1994; Wiskott, 2006). In the late 1940s, Pitts and McCulloch proposed neural networks to carry out both visual (translation, magnification) and auditory (melodic transposition) transformations (Pitts & McCulloch, 1947; McCulloch, 1951). Their representational model used diagonally crossing sets of projections on logarithmic retinotopic and cochleotopic cortical place maps to implement "shifter" circuits that would recognize angle and frequency ratios. However, if the underlying neural representations instead involve temporal patterns of spikes, then time-warping of these patterns, that is, stretching or compressing time intervals by a constant factor, can provide a general solution to the three auditory invariances (Boomsliter & Creel, 1962). The different temporal regimes associated with the three transformations would likely require processing at different levels. Fine-grained temporal information needed for recognizing harmonic ratios for recognitions of musical intervals and vowels is ubiquitous in early auditory stations, whereas coarse-grained temporal information for recognizing rhythmic patterns of events also exists over large portions of cerebral cortex (Thatcher & John, 1977).

13.4.5 Temporal Integration and Auditory Memory Mechanisms

Processing of sounds and sound sequences occurs over different time regimens that span echoic memory integration windows for pitch and timbre, and loudness summation, intermediate duration windows for melodic and rhythmic pattern integration, and still longer temporal windows for large-scale recurring patterns (Snyder, 2000, 2009; Trainor & Zatorre, 2009). When performing sequential matching tasks, human listeners can easily hold precise memories of pitch, timbre, loudness, location, and other auditory qualities for several seconds provided that subsequent distractions do not intervene (Demany & Semal, 2007). In musical contexts, tonal and rhythmic expectations can persist over even longer durations (Patel, 2008). To appreciate the complex interplay of multiple memory processes, one has only to think of an extended piece of tonal symphonic music, with its many excursions to and from tonal centers, metrical frames, and melodic motifs (Bigand, 1993).

The nature and locations of the various memory traces remain to be identified (Fritz et al., 2005), and their workings likely depend on the nature of the neural codes that are involved. For example, rate-place codes might entail persistently active subsets of neurons that encode particular features, whereas temporal codes might utilize reverberatory circuits that maintain temporal patterns of activity over time. Adaptation of neural responses over different timescales (ranging from milliseconds to several tens of minutes) likely plays an important role in the representation of temporal sound sequences in auditory cortex (Ulanovsky et al., 2004), and may potentially explain many aspects of music and speech perception.

13.4.6 Neural Requisites for Conscious Auditory Awareness

A great deal of progress has been made in the scientific study of the neural basis of consciousness over the last decade. The best current theories of the neural requisites of awareness involve the necessity of recurrent activation patterns for a given stimulus to become supraliminal (Lamme, 2006). Currently there is debate about whether recurrent corticocortical or thalamocortical activation of modality-specific pathways are sufficient (albeit without the ability for overt report), or whether recurrent activation patterns need also to include frontal and/or parietal regions as well. Recurrent activation of frontal regions results in systemic recurrence for support of global workspaces, while parietal activation of body/self maps may be essential for “ownership” of percepts (Pollen, 2008) or for providing a requisite level of attentional gain through associated basal ganglia circuits. Recurrent activation may facilitate attainment of a threshold degree of informational complexity (Tononi & Koch, 2008) or it may support dynamic regeneration of neuronal signals necessary for supporting sustained, stable systemic informational states in the first place (Cariani, 2000).

The vast majority of neurophysiological and psychophysical studies have involved visual experience, but any truly general theory of the neuronal basis for awareness needs to apply to other kinds of sensory experience as well. This makes the auditory system an ideal testing ground for theories developed using examples from vision. In the last decade striking auditory analogues to visual neglect syndromes and blindsight have been reported (Garde & Cowey, 2000, Clarke & Thiran, 2004). As in vision, it appears that body space representations in the parietal lobe must be engaged for auditory percepts to enter awareness, and also that the presence of auditory stimuli can be detected in the absence of direct experience of their qualities.

Many general and specific hypotheses concerning consciousness await investigation by auditory scientists. Is recurrent activation of frontal supramodal regions either essential or sufficient for auditory experience? Does conscious auditory awareness of an external sound event require completion of frontal–temporal feedback loop? Practically, to understand central tinnitus, one wants to identify the requisites for an endogenously generated neural pattern of activity to become part of conscious awareness. Beyond restoring auditory discriminatory capacities, it is also desirable to restore the subjective, felt texture of hearing in those who have lost or never had it, for example, the restoration of the experienced sound qualities of speech and music in cochlear implant users. Here a neurophenomenology that surveys the gamut of auditory experiences and identifies their neural correlates is a prerequisite. Whether in pursuit of restorative therapies or basic knowledge, auditory neuroscience will eventually develop such a neurophenomenological theory that will finally bridge the divide between our brains and our auditory experiences to provide useful and meaningful answers to fundamental questions of what and how we hear.

13.5 Summary

Although biological brains are impressively powerful informational engines, they are neither omnipotent nor infinitely complex—and there is no reason to believe that they cannot be understood by human minds properly equipped with the right conceptual tools. If the information functions of auditory cortex are to be understood, neurocomputational theories and neurophysiological experiments need to pay close attention to and strive to explain the large-scale structure of auditory perception and cognition. Because not all aspects of cortical structure and neural activity necessarily play critical roles in its informational functions, it is therefore essential that the cortical neural codes that do play such roles be identified as early as possible. As with the elucidation of the genetic code half a century ago, once the signals of the system are identified, understanding of the rest of the functional framework should quickly follow.

Acknowledgments This work was supported by research grants from the Air Force Office for Sponsored Research (FA9550-09-1-0119 to P. C.) and the National Institutes for Health (R01 DC05216 and R01 07657 to C. M.).

References

Abeles, M. (2003). Synfire chains. In M. A. Arbi (Ed.), *The handbook of brain theory and neural networks*, 2nd ed. (pp. 1143–1146). Cambridge, MA: MIT Press.

Abeles, M., Bergman, H., Margalit, E., & Vaadia, E. (1993). Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. *Journal of Neurophysiology*, 70, 1629–1638.

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Science*s, 8, 457–464.

Ando, Y., & Cariani, P. (2009). *Auditory and visual sensations*. New York: Springer.

Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. In W. A. Rosenbluth (Ed.), *Sensory communication* (pp. 217–234). Cambridge, MA: MIT Press.

von Bekesy, G. (1967). *Sensory inhibition*. Princeton, NJ: Princeton University Press.

Bendor, D., & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature*, 436, 1161–1165.

Bigand, E. (1993). Contributions of music to research on human auditory cognition. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 231–277). Oxford: Oxford University Press.

Bitterman, Y., Mukamel, R., Malach, R., Fried, I., & Nelken, I. (2008). Ultra-fine frequency tuning revealed in single neurons of human auditory cortex. *Nature*, 451, 197–201.

Bizley, J. K., & King, A. J. (2008). Visual-auditory spatial processing in auditory cortical neurons. *Brain Research*, 1242, 24–36.

Bizley, J. K., & Walker, K. M. M. (2010). Sensitivity and selectivity of neurons in the auditory cortex to the pitch, timbre, and location of sounds. *The Neuroscientist*, 16, 453–469.

Bizley, J. K., Walker K. M., King, A. J., & Schnupp, J. W. (2010). Neural ensemble codes for stimulus periodicity in auditory cortex. *Journal of Neuroscience*, 30(14), 5078–5091.

Boomsliter, P., & Creel, W. (1962). The long pattern hypothesis in harmony and hearing. *Journal of Music Theory*, 5, 2–31.

Boring, E. G. (1942). *Sensation and perception in the history of experimental psychology*. New York: Appleton-Century-Crofts.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

Brugge, J. F., Reale, R. A., & Hind, J. E. (1996). The structure of spatial receptive fields of neurons in primary auditory cortex of the cat. *Journal of Neuroscience*, 16, 4420–4437.

Bullock, T. H. (1992). Introduction to induced rhythms: A widespread heterogeneous class of oscillations. In E. Basar & T. H. Bullock (Eds.), *Induced rhythms in the brain* (pp. 1–26). Boston: Birkhauser.

Buzsáki, G. (2006). *Rhythms of the brain*. New York: Oxford University Press.

Cariani, P. (1995). As if time really mattered: Temporal strategies for neural coding of sensory information. *Communication and Cognition–Artificial Intelligence (CC-AI)*, 12, 161–229. Reprinted in K. Pribram (Ed.), *Origins: Brain and self-organization* (pp. 208–252). Hillsdale, NJ: Lawrence Erlbaum.

Cariani, P. (1999). Temporal coding of periodicity pitch in the auditory system: An overview. *Neural Plasticity*, 6, 147–172.

Cariani, P. (2000). Regenerative process in life and mind. In J. L. R. Chandler & G. Van de Vijver (Eds.), *Closure: Emergent organizations and their dynamics*. *Annals of the New York Academy of Sciences*, 901, 26–34.

Cariani, P. (2001a). Neural timing nets. *Neural Networks*, 14, 737–753.

Cariani, P. (2001b). Temporal coding of sensory information in the brain. *Acoustic Science & Technology*, 22, 77–84.

Cariani, P. (2002). Temporal codes, timing nets, and music perception. *Journal of New Music Research*, 30, 107–136.

Cariani, P. (2004). Temporal codes and computations for sensory representation and scene analysis. *IEEE Transactions on Neural Networks*, *Special Issue on Temporal Coding for Neural Information Processing*, 15, 1100–1111.

Cariani, P., & Delgutte, B. (1996). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. II. Pitch shift, pitch ambiguity, phase-invariance, pitch circularity, and the dominance region for pitch. *Journal of Neurophysiology*, 76, 1698–1734.

Carr, C. E. (1993). Processing of temporal information in the brain. *Annual Review of Neuroscience*, 16, 223–243.

Chait, M., Poeppel, D., de Cheveigne, A., & Simon, J. Z. (2007). Processing asymmetry of transitions between order and disorder in human auditory cortex. *Journal of Neuroscience*, 27(19), 5207–5214.

Chase, S. M., & Young, E. D. (2006). Spike-timing codes enhance the representation of multiple simultaneous sound-localization cues in the inferior colliculus. *Journal of Neuroscience*, 26, 3889–3898.

Chechik, G., Anderson, M. J., Bar-Yosef, O., Young, E. D., Tishby, N., & Nelken, I. (2006). I. Reduction of information redundancy in the ascending auditory pathway. *Neuron*, 51, 359–368.

Cherry, C. (1961). Two ears–but one world. In W. A. Rosenblith (Ed.). *Sensory communication* (pp. 99–117). New York: MIT Press/John Wiley & Sons.

Cherry, C. (1966). *On human communication*. Cambridge, MA: MIT Press.

de Cheveigné, A. (2005). Pitch perception models. In C. J. Plack, A. J. Oxenham, R. Fay & A. N. Popper (Eds.), *Pitch: Neural coding and perception* (pp. 169–233). New York: Springer.

Chow, K. L. (1951). Numerical estimates of the auditory central nervous system of the rhesus monkey. *Journal of Comparative Neurology*, 95, 159–175.

Clarke, S., & Thiran, A. B. (2004) Auditory neglect: What and where in auditory space. *Cortex*, 40(2), 291–300.

Damasio, A. R., & Damasio, H. (1994). Cortical systems for retrieval of concrete knowledge: The convergence zone framework. In C. Koch & J. L. Davis (Eds.), *Large-scale neuronal theories of the brain* (pp. 61–74), Cambridge, MA: MIT Press.

Delgutte, B., & Kiang, N. Y. S. (1984). Speech coding in the auditory nerve: I. Vowel-like sounds. *Journal of the Acoustical Society of America*, 75(3), 866–878.

Demany, L., & Semal, C. (2007). The role of memory in auditory perception. In W. A. Yost, A. N. Popper, & R. R. Fay (Eds.), *Auditory perception of sound sources* (pp. 77–113). New York: Springer.

DeWeese, M. R., Hromadka, T., & Zador, A. M. (2005). Reliability and representational bandwidth in the auditory cortex. *Neuron*, 48, 479–488.

Eggermont, J. J. (1990). *The correlative brain: Theory and experiment in neural interaction*. Berlin: Springer.

Emmers, R. (1981). *Pain: A spike-interval coded message in the brain*. New York: Raven Press.

Evans, E. F., Pratt, S. R., Spenner, H., & Cooper, N. P. (1992). Comparisons of physiological and behavioural properties: Auditory frequency selectivity. In Y. Cazals, K. Horner, & L. Demany (Eds.), *Auditory physiology and perception*, Vol. 83 (pp. 159–169). Oxford: Pergamon.

Fishman, Y. I., Reser, D. H., Arezzo, J. C., & Steinschneider, M. (2000). Complex tone processing in primary auditory cortex of the awake monkey. I. Neural ensemble correlates of roughness. *Journal of the Acoustical Society of America*, 108, 235–246.

Fritz, J., Shamma, S., Elhilali, M., & Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nature Neuroscience*, 6, 1216–1223.

Fritz, J., Mishkin, M., & Saunders, R. C. (2005). In search of an auditory engram. *Proceedings of the National Academy of Sciences of the USA*, 102, 9359–9364.

6–Furukawa, S., Xu, L., & Middlebrooks, J. C. (2000). Coding of sound-source location by ensembles of cortical neurons. *Journal of Neuroscience*, 20, 1216–1228.

Garde, M. M., & Cowey, A. (2000). "Deaf hearing": Unacknowledged detection of auditory stimuli in a patient with cerebral deafness. *Cortex*, 36(1), 71–79.

Greene, P. H. (1962). On looking for neural networks and "cell assemblies" that underlie behavior. I. Mathematical model. II. Neural realization of a mathematical model. *Bulletin of Mathematical Biophysics*, 24, 247–275, 395–411.

Handel, S. (1989). *Listening: An introduction to the perception of auditory events*. Cambridge, MA: MIT Press.

Heil, P. (1997). Auditory cortical onset responses revisited. I. First-spike timing. *Journal of Neurophysiology*, 77, 2616–2641.

Hromadka, T., DeWeese, M. R., & Zador, A. M. (2008). Sparse representation of sounds in the unanesthetized auditory cortex. *Public Library of Science (PLoS) Biology*, 6, e16.

Huetz, C., Gourevitch, B., & Edeline, J. M. (2011). Neural codes in the thalamocortical auditory system: From artificial stimuli to communication sounds. *Hearing Research*, 271(1–2), 147–158.

Huron, D. B. (2006). *Sweet anticipation: Music and the psychology of expectation*. Cambridge, MA: MIT Press.

Izhikevich, E. M. (2006). Polychronization: Computation with spikes. *Neural Computation*, 18, 245–282.

Jeffress, L. A. (1948). A place theory of sound localization. *Journal of Comparative and Physiological Psychology*, 41, 35–39.

Kayser, C., Petkov, C. I., & Logothetis, N. K. (2008). Visual modulation of neurons in auditory cortex. *Cerebral Cortex*, 18, 1560–1574.

Koch, C. (2004). *The quest for consciousness: A neurobiological approach*. Denver: Roberts & Co.

Köhler, W. (1947). *Gestalt psychology: An introduction to new concepts in modern psychology*. New York: Liveright.

Kowalski, N., Depireux, D. A., & Shamma, S. A. (1996). Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra. *Journal of Neurophysiology*, 76, 3503–3523.

Lamme, V. A. (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Sciences*, 10, 494–501.

Laurent, G. (2006). Shall we even understand the fly's brain? In J. L. van Hemmen & T. J. Sejnowski (Eds.), *23 problems in systems neuroscience* (pp. 3–21). Oxford: Oxford University Press.

Lee, K. M., Skoe, E., Kraus, N., & Ashley, R. (2009). Selective subcortical enhancement of musical intervals in musicians. *Journal of Neuroscience*, 29, 5832–5840.

Licklider, J. C. R. (1959). Three auditory theories. In S. Koch (Ed.), *Psychology: A study of a science. Study I. Conceptual and systematic*, Vol. I: *Sensory*, *perceptual*, *and physiological formulations* (pp. 41–144). New York: McGraw-Hill.

von der Malsberg, C. (1994). The correlation theory of brain function. In E. Domany, J. L. van Hemmen, & K. Schulten (Eds.), *Models of neural networks II: Temporal aspects of coding and information processing in biological systems* (pp. 95–120). New York: Springer.

Marr, D. (1982). *Vision: A computational approach*. San Francisco: Freeman & Co.

McAdams, S., & Giordano, B. L. (2009). The perception of musical timbre. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (pp. 72–80). Oxford: Oxford University Press.

McCulloch, W. S. (1951). Why the mind is in the head. In L. A. Jeffress (Ed.), *Cerebral mechanisms of behavior* (pp. 42–111). New York: John Wiley & Sons.

McDermott, J., & Oxenham, A. (2008) Music perception, pitch, and the auditory system. *Current Opinion in Neurobiology*, 18, 452–463.

Mesulam, M. M. (2000). *Principles of behavioral and cognitive neurology*. New York: Oxford University Press.

Miller, L. M., Escabi, M. A., Read, H. L., & Schreiner, C. E. (2001). Functional convergence of response properties in the auditory thalamocortical system. *Neuron*, 32, 151–160.

Miller, L. M., Escabí, M. A., Read, H. L., & Schreiner, C. E. (2002). Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. *Journal of Neurophysiology*, 87, 516–527.

Monahan, P. J., & Idsardi, W. J. (2010). Auditory sensitivity to formant ratios: Toward an account of vowel normalization. *Language and Cognitive Processes*, 25(6), 808–839.

Moore, B. C. (1973). Frequency difference limens for short-duration tones. *Journal of the Acoustical Society of America*, 54, 610–619.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, 118, 2544–2590.

Nelken, I., Fishbach, A., Las, L., Ulanovsky, N., & Farkas, D. (2003). Primary auditory cortex of cats: Feature detection or something else? *Biological Cybernetics*, 89, 397–406.

Palmer, A. R. (1992). Segregation of the responses to paired vowels in the auditory nerve of the guinea pig using autocorrelation. In M. E. H. Schouten (Ed.), *The auditory processing of speech* (pp. 115–124). Berlin: Mouton de Gruyter.

Panzeri, S., Brunel, N., Logothetis, N. K., & Kayser, C. (2009). Sensory neural codes using multiplexed temporal scales. *Trends in Neuroscience*s, 33(3), 111–120.

Patel, A. D. (2008). *Music*, *language and the brain*. Oxford: Oxford University Press.

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, 36, 767–776.

Penagos, H., Melcher, J. R., & Oxenham, A. J. (2004). A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *Journal of Neuroscience*, 24, 6810–6815.

Perkell, D. H., & Bullock T. H. (1968). Neural coding. *Neurosciences Research Program Bulletin*, 6, 221–348.

Phillips, D. P., Semple, M. N., Calford, M. B., & Kitzes, L. M. (1994). Level-dependent representation of stimulus frequency in cat primary auditory cortex. *Experimental Brain Research*, 102, 210–226.

Phillips, D. P., Hall, S. E., & Boehnke, S. E. (2002). Central auditory onset responses, and temporal asymmetries in auditory perception. *Hearing Research*, 167, 192–205.

Reprinted in W. S. McCulloch (Ed.), *Embodiments of mind* (pp. 46–66). Cambridge, MA: MIT Press, 1965.Pitts, W., & McCulloch, W. S. (1947). How we know universals: The perception of auditory and visual forms. *Bulletin of Mathematical Biophysics*, 9, 127–147. Reprinted in W. S. McCulloch (Ed.), *Embodiments of mind* (pp. 46–66). Cambridge, MA: MIT Press, 1965.

Pollen, D. A. (2008). Fundamental requirements for primary visual perception. *Cerebral Cortex*, 18, 1991–1998.

Rauschecker, J. P., & Tian, B. (2004). Processing of band-passed noise in the lateral auditory belt cortex of the rhesus monkey. *Journal of Neurophysiology*, 91, 2578–2589.

Rieke, F., Warland, D., de Ruyter van Steveninck, R., & Bialek, W. (1997). *Spikes: Exploring the neural code*. Cambridge, MA: MIT Press.

Sadagopan, S., & Wang, X. (2008). Level invariant representation of sounds by populations of neurons in primary auditory cortex. *Journal of Neuroscience*, 28(13), 3415–3426.

Schreiner, C. E., & Langner, G. (1988). Coding of temporal patterns in the central auditory nervous system. In G. Edelman (Ed.), *Auditory function: Neurobiological bases of hearing* (pp. 337–361). New York: John Wiley & Sons.

Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, 32(1), 9–17.

Shamma, S. A., & Micheyl, C. (2010). Behind the scenes of auditory perception. *Current Opinion in Neurobiology*, 20, 361–366.

Shamma, S. A., Elhilali, M., & Micheyl, C. (2010). Temporal coherence and attention in auditory scene analysis. *Trends in Neurosciences*, 34(3), 114–123.

1–Singer, W. (2003). Synchronization, binding, and expectancy. In M. A. Arbib (Ed.). *The handbook of brain theory and neural networks*, 2nd ed. (pp. 1136–1143). Cambridge, MA: MIT Press.

Snyder, B. (2000). *Music and memory*. Cambridge, MA: MIT Press.

Snyder, B. (2009). Memory for music. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (pp. 171–183). Oxford: Oxford University Press.

Thatcher, R. W., & John, E. R. (1977). *Foundations of cognitive processes*. Hillsdale, NJ: Lawrence Erlbaum.

Tian, B., & Rauschecker, J. P. (1994). Processing of frequency-modulated sounds in the cat's anterior auditory field. *Journal of Neurophysiology*, 71, 1959–1975.

Tononi, G., & Koch, C. (2008). The neural correlates of consciousness: An update. *Annals of the New York Academy of Sciences*, 1124, 239–61.

Trainor, L. J., & Zatorre, R. J. (2009). The neurobiological basis of musical expectations. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (pp. 171–183). Oxford: Oxford University Press.

Tramo, M. J., Cariani, P. A., Koh, C. K., Makris, N., & Braida, L. D. (2005). Neurophysiology and neuroanatomy of pitch perception: Auditory cortex. *Annals of the New York Academy of Sciences*, 1060, 148–174.

Tzounopoulos, T., Kim, Y., Oertel, D., & Trussell, L. O. (2004). Cell-specific, spike timing-dependent plasticities in the dorsal cochlear nucleus. *Nature Neuroscience*, 7, 719–725.

Ulanovsky, N., Las, L., Farkas, D., & Nelken, I. (2004). Multiple time scales of adaptation in auditory cortex neurons. *Journal of Neuroscience*, 24, 10440–10453.

Van Rullen, R., & Thorpe, S. J. (2001). Rate coding versus temporal order coding: What the retinal ganglion cells tell the visual cortex. *Neural Computation*, 13, 1255–1283.

Villa, A. E. (2000). Empirical evidence about temporal structure in multi-unit recordings. In R. Miller (Ed.), *Time and the brain* (pp. 1–52). Amsterdam: Harwood.

Wallace, M. N., Shackleton, T. M., & Palmer, A. R. (2002). Phase-locked responses to pure tones in the primary auditory cortex. *Hearing Research*, 172, 160–171.

Wang, D. (2002). *The time dimension for neural computation* (pp. 1–40). Columbus, OH: Center for Cognitive Science and the Department of Computer & Information Science, The Ohio State University.

Wang, X. (2007). Neural coding strategies in auditory cortex. *Hearing Research*, 229, 81–93.

Warren, J. D., Uppenkamp, S., Patterson, R. D., & Griffiths, T. D. (2003). Separating pitch chroma and pitch height in the human brain. *Proceedings of the National Academy of Sciences of the USA*, 100, 10038–10042.

Winer, J. A. (1992). The functional architecture of the medial geniculate body and the primary auditory cortex. In D. B. Webster, A. N. Popper, & R. R. Fay (Eds.), *The mammalian auditory pathway: Neuroanatomy* (pp. 222–286). New York: Springer.

Winkler, I., Denham, S. L., & Nelken, I. (2009). Modeling the auditory scene: Predictive regularity representations and perceptual objects. *Trends in Cognitive Sciences*, 13, 532–540.

Wiskott, L. (2006). How does our visual system achieve shift invariance? In J. L. van Hemmen (Ed.), *23 problems in systems neuroscience* (pp. 322–340). Oxford: Oxford University Press.

Index

D. Poeppel et al. (eds.), *The Human Auditory Cortex*, Springer Handbook of Auditory Research 43, DOI 10.1007/978-1-4614-2314-0,