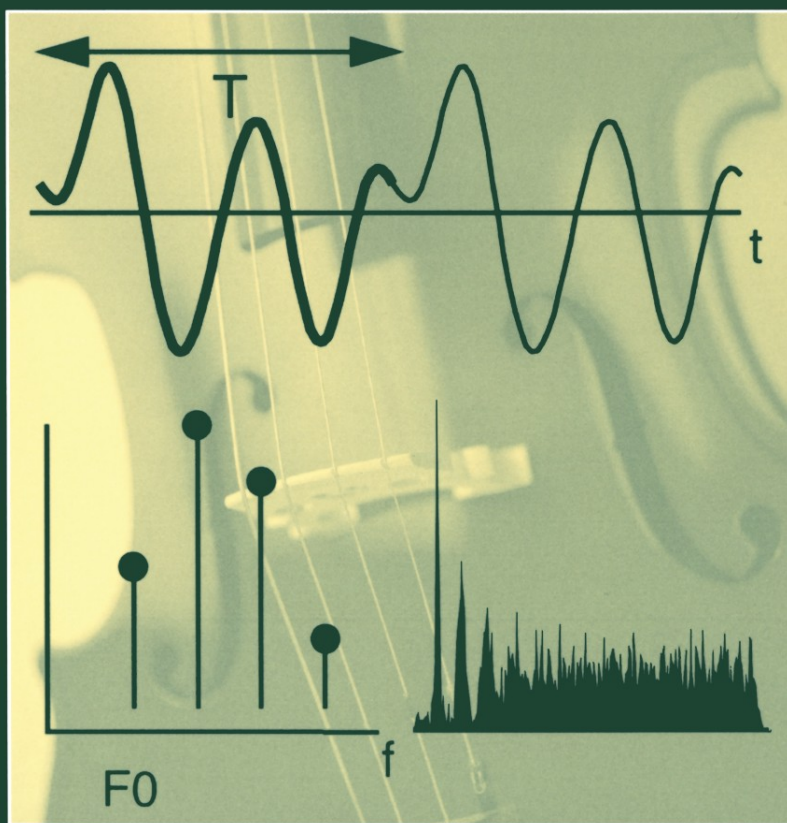


**Christopher J. Plack**  
**Andrew J. Oxenham**  
**Richard R. Fay**  
**Arthur N. Popper** Editors

# Pitch

## Neural Coding and Perception



SPRINGER HANDBOOK OF  
AUDITORY RESEARCH

---

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

Christopher J. Plack  
Andrew J. Oxenham  
Richard R. Fay  
Arthur N. Popper  
Editors

# Pitch

## Neural Coding and Perception

With 74 illustrations and 5 color illustrations

Christopher J. Plack  
Department of Psychology  
University of Essex  
Colchester CO4 3SQ  
United Kingdom  
cplack@essex.ac.uk

Andrew J. Oxenham  
Research Laboratory of Electronics  
Massachusetts Institute of Technology  
Cambridge, MA 02139, USA  
oxenham@mit.edu

Richard R. Fay  
Parmly Hearing Institute and  
Department of Psychology  
Loyola University of Chicago  
Chicago, IL 60626, USA  
rfay@wpo.it.luc.edu

Arthur N. Popper  
Department of Biology  
University of Maryland  
College Park, MD 20742, USA  
apopper@umd.edu

*Cover illustration:* The image includes parts of Figures 4.6 and 6.4 appearing in the text.

Library of Congress Cataloging-in-Publication Data

Pitch: neural coding and perception / [edited by] Christopher J. Plack, Andrew J. Oxenham, Richard R. Fay, Arthur N. Popper.

p. cm.—(Springer handbook of auditory research; v. 24)

Includes bibliographical references and index.

ISBN 10: 0-387-23472-1 (alk. paper)

1. Auditory perception. 2. Musical pitch. I. Plack, Christopher J. II. Oxenham, Andrew J. III. Fay, Richard R. IV. Series.

QP465.P545 2005

152.1'52—dc22

2004057843

ISBN 10: 0-387-23472-1

Printed on acid-free paper

ISBN 13: 978-0387-23472-4

© 2005 Springer Science+Business Media, Inc.

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, Inc., 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden. The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed in the United States of America. (EB)

9 8 7 6 5 4 3 1

springeronline.com



Each of the editors takes pleasure in dedicating this volume  
to his parents in gratitude for their support and guidance:

Audrey and Jim Plack  
Margaret and John Oxenham  
Ingrid and Charles Fay  
Evelyn and Martin Popper

# 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, postdoctoral 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.

RICHARD R. FAY, Chicago, Illinois  
ARTHUR N. POPPER, College Park, Maryland

# Volume Preface

The seeds for this volume on pitch were sown in October 2001, when Wolfgang Stenzel, Andrew Oxenham, and Chris Plack met for dinner in a Spanish restaurant in Bremen, Germany. They discussed the possibility of organizing a conference on pitch perception to be hosted by the Hanse Wissenschaftskolleg (Hanse Institute for Advanced Study) in Delmenhorst (Wolfgang Stenzel administers the Neurosciences and Cognitive Sciences Program at the Institute). The proposal to the Institute began as follows: “Although pitch has been considered an important area of auditory research since the nineteenth century, some of the most significant developments in our understanding of this phenomenon have occurred comparatively recently. The time is ripe for a meeting that brings together experts from several different disciplines to share ideas and gain insights into the fundamental (and still largely unsolved) problem of how the brain processes the pitch of acoustic stimuli.” The conference took place August 2002, bringing together scientists in the fields of neuroscience, computational modeling, cognitive science, and music psychology.

Rather than publish a standard conference proceedings, Plack and Oxenham approached Arthur Popper and Richard Fay about producing this volume, which is a “stand-alone” review of the current state of pitch research, inspired by (but not limited to) the presentations and discussions at the conference. All the chapter authors attended the conference, and, like the conference, the volume brings together researchers from a range of different disciplines. It is hoped that the reader may obtain a broad view of the topic from basic neurophysiology to more cognitive processes.

Chapter 1, by Plack and Oxenham, provides a definition of pitch and an overview of the field. A description of the basic psychophysics of pitch is the focus of Chapters 2 and 3. Plack and Oxenham (Chapter 2) describe how human perceptions are related to the physical characteristics of the stimulus and a similar approach is taken in a discussion of psychophysical studies on nonhuman animals by Shofner in Chapter 3. In Chapter 4, Winter examines in detail the neural representation of periodicity information and describes how and where in the auditory system periodicity information may be processed and extracted. Animal experiments are required for a detailed investigation of neural mecha-

nisms. However, it is also possible to observe more general physiological processes in the *human* auditory system. In Chapter 5, Griffiths explains how modern brain-imaging techniques (PET, fMRI, EEG, and MEG) have enabled researchers to probe the regions responsible for pitch processing in the human brain. In Chapter 6, de Cheveigné provides a detailed taxonomy of pitch models using a rich historical and conceptual context. He highlights the commonalities between models and outlines the bases for selecting between them. Pitch perception for listeners with hearing impairment and with cochlear implants is discussed in Chapter 7 by Moore and Carlyon. In addition to the clinical benefits, such as the design of prostheses, readers of this chapter will be aware of just how much we can learn about “normal” pitch mechanisms by examining the consequences of disrupted auditory processing.

In Chapter 8, Darwin considers one of the most important uses of periodicity information, the segregation of sounds from different sources and the grouping of frequency components from the same source. Finally, in Chapter 9, Bigand and Tillmann consider what may be regarded as “higher-level” or more cognitive aspects of pitch perception, with particular reference to the perception of music.

The chapter topics of this volume have been discussed more briefly and from other viewpoints in other volumes of the *Springer Handbook of Auditory Research* series. The psychoacoustics of spectral, temporal, and pitch processing have been presented earlier in Volume 3 (*Human Psychophysics*). Comparative studies of hearing at the anatomical, physiological, and behavioral levels have been extensively treated in Volumes 4 (*Comparative Hearing: Mammals*), 11 (*Comparative Hearing: Fish and Amphibians*), and 13 (*Comparative Hearing: Birds and Reptiles*). Neurophysiological studies of coding and auditory representations relevant to pitch perception have been discussed in Volumes 4, 11, and 13 and in Volumes 2 (*The Mammalian Auditory Pathway: Neurophysiology*) and 15 (*Integrative Functions in the Mammalian Auditory Pathway*). Models of auditory information processing, including pitch, were introduced in Volume 8 (*The Cochlea*) and more extensively developed in Volume 6 (*Auditory Computation*). More information on pitch perception and the hearing functions of persons with hearing impairments and cochlear implants can be found in Volumes 7 (*Clinical Aspects of Hearing*) and 20 (*Cochlear Implants: Auditory Prostheses and Electric Hearing*).

We thank the authors of the chapters for giving so much of their time to the project and for enduring the nagging of the editors. We hope that you agree that the scholarship exhibited is of the highest standard. The volume would not be what it is without our quality-control team of expert chapter reviewers: Josh Bernstein, John Culling, Alain de Cheveigné, Steve McAdams, Christophe Micheyl, Brian Moore, Alan Palmer, Daniel Pressnitzer, and Lutz Wiegand. For no reward, these noble individuals made detailed and constructive comments on earlier drafts, excising the chaff and invigorating the wheat. We would also like to express our appreciation to the staff at Springer, particularly Janet Slobodien. Finally, we thank the Hanse Wissenschaftskolleg for facilitating the conference

that led to this book and for providing financial support in covering the additional cost of the color figures in this volume.

CHRISTOPHER J. PLACK, Colchester, United Kingdom  
ANDREW J. OXENHAM, Cambridge, Massachusetts  
RICHARD R. FAY, Chicago, Illinois  
ARTHUR N. POPPER, College Park, Maryland

# Contents

|                                                                                                                  |     |
|------------------------------------------------------------------------------------------------------------------|-----|
| Series Preface . . . . .                                                                                         | vii |
| Volume Preface . . . . .                                                                                         | ix  |
| Contributors . . . . .                                                                                           | xv  |
| Chapter 1    Overview: The Present and Future of Pitch . . . . .                                                 | 1   |
| CHRISTOPHER J. PLACK AND ANDREW J. OXENHAM                                                                       |     |
| Chapter 2    The Psychophysics of Pitch . . . . .                                                                | 7   |
| CHRISTOPHER J. PLACK AND ANDREW J. OXENHAM                                                                       |     |
| Chapter 3    Comparative Aspects of Pitch Perception . . . . .                                                   | 56  |
| WILLIAM P. SHOFNER                                                                                               |     |
| Chapter 4    The Neurophysiology of Pitch . . . . .                                                              | 99  |
| IAN M. WINTER                                                                                                    |     |
| Chapter 5    Functional Imaging of Pitch Processing . . . . .                                                    | 147 |
| TIMOTHY D. GRIFFITHS                                                                                             |     |
| Chapter 6    Pitch Perception Models . . . . .                                                                   | 169 |
| ALAIN DE CHEVEIGNÉ                                                                                               |     |
| Chapter 7    Perception of Pitch by People with Cochlear Hearing<br>Loss and by Cochlear Implant Users . . . . . | 234 |
| BRIAN C.J. MOORE AND ROBERT P. CARLYON                                                                           |     |
| Chapter 8    Pitch and Auditory Grouping . . . . .                                                               | 278 |
| CHRISTOPHER J. DARWIN                                                                                            |     |
| Chapter 9    Effect of Context on the Perception of Pitch Structures . .                                         | 306 |
| EMMANUEL BIGAND AND BARBARA TILLMANN                                                                             |     |
| Index . . . . .                                                                                                  | 353 |

# Contributors

EMMANUEL BIGAND

L.E.A.D.-C.N.R.S. UMR 5022, Université de Bourgogne, F-21000 Dijon, France

ROBERT P. CARLYON

MRC Cognition and Brain Sciences Unit, Cambridge CB2 2EF, United Kingdom

CHRISTOPHER J. DARWIN

Department of Psychology, University of Sussex, Brighton BN1 9QG, United Kingdom

ALAIN DE CHEVEIGNÉ

CNRS/IRCAM, 75004 Paris, France

TIMOTHY D. GRIFFITHS

Auditory Group, Newcastle University Medical School, Newcastle NE2 4HH, United Kingdom

BRIAN C.J. MOORE

Department of Psychology, University of Cambridge, Cambridge CB2 3EB, United Kingdom

ANDREW J. OXENHAM

Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

CHRISTOPHER J. PLACK

Department of Psychology, University of Essex, Colchester CO4 3SQ, United Kingdom

WILLIAM P. SHOFNER

Parmly Hearing Institute, Loyola University of Chicago, Chicago, IL 60626,  
USA

BARBARA TILLMANN

CNRS UMR 5020 Neurosciences et Systèmes Sensoriels, F69366 Lyon Cedex  
07, France

IAN M. WINTER

The Physiological Laboratory, University of Cambridge, Cambridge CB2 3EG,  
United Kingdom



# Overview: The Present and Future of Pitch

CHRISTOPHER J. PLACK AND ANDREW J. OXENHAM

## 1. Definition of Pitch

This book is about pitch, so our first duty is to define exactly what we mean by the word. Unfortunately this is not a straightforward exercise, as many different definitions have been proposed over the years. The definitions fall into two broad categories: those that make a reference to the association between pitch and the musical scale and those that avoid a reference to music.

### *1.1 Definitions Referring to Music*

In 1960, the American Standards Association was explicit about the relationship between pitch and music, defining pitch as “that attribute of auditory sensation in terms of which sounds may be ordered on a musical scale” (ASA 1960). An important aspect of this definition is that pitch is an attribute of *sensation*. The word “pitch” should not be used to refer to a *physical* attribute of a sound.

Some authors have used the ability of a sound to produce recognizable musical melodies (by varying repetition rate, modulation rate, etc.) as a test of whether that sound evokes a pitch (e.g., Burns and Viemeister 1976). Put another way, pitch is the perceptual attribute of a sound that can be used to produce melodies. Although most would agree that the production of melodies is *sufficient* to prove that a sound can evoke a pitch, some would not regard it as a *necessary* condition.

### *1.2 Definitions Not Referring to Music*

The more recent American National Standards definition dispenses with the musical reference: “Pitch [is] that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high. Pitch depends primarily on the frequency content of the sound stimulus, but it also depends on the sound pressure and the waveform of the stimulus” (ANSI 1994). This appears to be a fairly broad definition, requiring the words “low” and

“high” to be associated with pitch or frequency, rather than with loudness or intensity, for example. Also, this definition seems to include what some would regard as timbral effects, such as the increase in the “brightness” of a sound as the level of its high-frequency content increases.

It is also possible to have an operational definition that does not depend on music, based on a pure-tone reference. For example: “A sound can be said to have a certain pitch if it can be reliably matched by adjusting the frequency of a pure tone of arbitrary amplitude” (Hartmann 1997). In this definition, it has to be assumed that the listener is matching on the basis of pitch, rather than loudness or timbre, or perhaps some combination of all three. The definition includes stimuli that cannot be used to produce recognizable melodies, for example, pure tones with frequencies above 5000 Hz.

### 1.3 Conclusion

The definitions cited in this section are a small, but representative, sample of the number of different definitions of pitch that can be found in the literature. For the purposes of this book we decided to take a conservative approach, and to focus on the relationship between pitch and musical melodies. Following the earlier ASA definition, we define pitch as “that attribute of sensation whose variation is associated with musical melodies.” Although some might find this too restrictive, an advantage of this definition is that it provides a clear procedure for testing whether or not a stimulus evokes a pitch, and a clear limitation on the range of stimuli that we need to consider in our discussions.

## 2. Why Is Pitch Important?

Many of the sounds in our environment have acoustic waveforms that repeat over time. These sounds are often perceived as having a pitch that corresponds to the repetition rate of the sound. Vowel sounds in speech are “voiced” and can be associated with a pitch. Many musical instruments produce a pitch that enables them to produce melodies and chords. More generally, sound sources in our environment often have characteristic rates of vibration: for example, the high-frequency ringing of a drinking glass, or the varying low-frequency revolutions of a car engine.

Pitch is an important attribute of any auditory stimulus in and of itself. It is arguably the most relevant perceptual dimension in most forms of Western music and is also important for speech communication, carrying important prosody information in languages such as English, but also carrying semantic information in tone languages, such as Mandarin. As discussed by Darwin (Chapter 8), however, another very important aspect of pitch is that it enhances our ability to perceptually segregate sound sources, based on differences in fundamental frequency (F0). Pitch can also be used to *group together* the individual sound components, or harmonics, that arise from the same vibrating source. Pitch is

therefore of primary importance in defining and differentiating our acoustic environment.

Understanding pitch perception is not a purely academic exercise. As our understanding of how the human auditory system processes pitch increases, so too will our ability to harness the findings in a variety of applications, such as speech recognition systems that are more robust to interfering sounds; so far, even the best technical systems fail miserably in distinguishing different acoustic sources, when compared to the performance of the human auditory system. Understanding pitch mechanisms will also help us to design prostheses, such as hearing aids and cochlear implants, that maximize the relevant information that is available to the listener.

### 3. Summary of Findings and Future Directions

Research into pitch perception has generated such a diverse range of findings that it is difficult, and perhaps unfair, to summarize these results in just a few paragraphs. We will start with what we think we know, move on to the more controversial aspects, and finish with a list of some unresolved issues and speculate on how they may be resolved.

We know that, with the exception of pure tones, pitch is not a simple function of the spectral content of a sound (Plack and Oxenham, Chapter 2; Shofner, Chapter 3; Moore and Carlyon, Chapter 7). Rather, pitch is related more closely to the repetition rate (or in some cases envelope repetition rate) of the sound, with a range from about 30 Hz to 5000 Hz. Sounds with the same repetition rate and very different spectra often have the same pitch (e.g., a pure tone with a frequency of 100 Hz and a complex tone with high harmonics with an F0 of 100 Hz), and sounds with similar spectra can have very different pitches (e.g., a wideband noise amplitude modulated at 100 or 200 Hz). This means that the frequency to place mapping performed by the cochlea does not equate to a frequency to *pitch* mapping. The auditory system combines information *across* cochlear location in order to derive the pitch of some stimuli (complex tones with low numbered harmonics). Furthermore, although pitch may be represented partly in terms of the gross activity of different regions of the cochlea, it is almost certainly represented in terms of the precise timing of neural impulses in the auditory nerve and at higher centers in the auditory system. Two stimuli with no perceptible spectral differences can produce very different pitches.

Physiological measurements (Winter, Chapter 4) and simulations using computational models (de Cheveigné, Chapter 6) have demonstrated that the repetition rates of stimuli are very well represented by the pattern of phase locking in the auditory nerve. For many researchers, the classical arguments of place versus time have been replaced by arguments about how and where in the auditory pathway the phase-locked activity is analyzed. The maximum frequency to which a fiber will phase lock declines from the auditory nerve to the auditory

cortex and it is thought that somewhere in the brainstem, possibly in the cochlear nucleus and/or the inferior colliculus, the synchrony representation is converted into a rate-place representation, in which different neurons code for different pitches in terms of overall firing rate. The existence of pitches arising from the detection of variations in binaural correlation suggests that at least some of these “pitch neurons” must be linked to binaural mechanisms.

Before we get carried away, however, we should consider a few unpleasant complications to this story. First, there is some evidence, not conclusive admittedly, suggesting that there are separate pitch mechanisms for stimuli with low harmonics that are resolved by the cochlea and for stimuli with high harmonics that are not resolved by the cochlea. There has been a recent resurgence in the old idea that there may be pitch templates for the resolved harmonics, with slots at harmonic intervals. One possibility is that an individual template neuron, tuned to a particular pitch, may receive input from neurons responding to information at specific harmonic frequencies. The individual frequencies converging on a template may be derived either from the spatial cochlear representation (rate-place) or possibly from a temporal analysis of the phase-locked response to each harmonic. For the *unresolved* harmonics the picture is murkier still, with some evidence that the gross rate of envelope fluctuations may have a greater influence on pitch than the precise timing of envelope peaks, a finding at odds with models of pitch based on the detection of temporal regularity.

Experiments on auditory grouping have contributed to our understanding of higher-level (cortical?) processes, and they also have important implications for our understanding of basic auditory mechanisms (Darwin, Chapter 8). F0 and harmonicity are important cues for the grouping and segregation of simultaneous and sequential sound components, and conversely grouping mechanisms determine which components contribute to the pitch that is heard. For example, the finding that the contribution of individual harmonics to the pitch of a complex tone can be influenced by sounds before and after the complex (e.g., a sequence of pure tones at a harmonic frequency) suggests that there is a considerable top-down influence on the pitch mechanism, so that the inclusion of frequency components into the analysis is governed partly by long-term, high-level processes.

Finally, we move on to the issue of how the extracted pitch is used to identify auditory objects and patterns, particularly with regard to speech and music. Imaging studies suggest that such processing may occur in the temporal and frontal lobes (Griffiths, Chapter 5), and probably involves the interaction of billions of neurons. Although we may never be able to understand these processes at the level of individual neurons, results of experiments on high-level perception, such as those described by Bigand and Tillmann (Chapter 9), allow an understanding at a different level of explanation. As with many perceptual phenomena, the sensation produced by a pitch or pitches is heavily dependent on the acoustic context and on prior experience, again implying that top-down processes are working at this level of analysis.

Figure 1.1 is a schematic (and simplistic) illustration of how the main processing stages and neural representations in pitch perception *might* be organized.

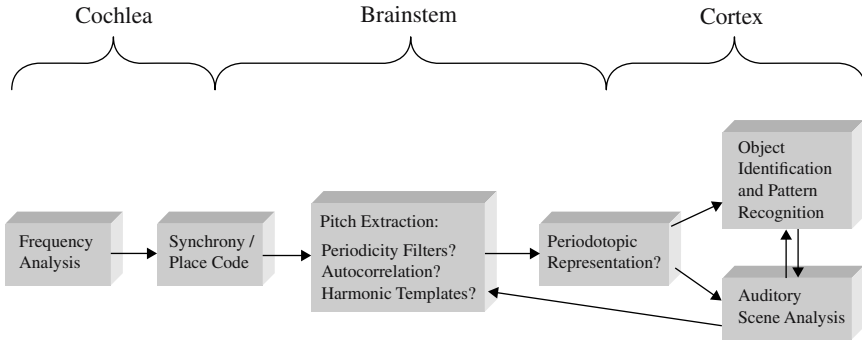


FIGURE 1.1. A crude illustration of how and where pitch might be processed in the auditory system.

The preceding discussion has highlighted huge gaps in our knowledge regarding the underlying mechanisms. Some of the fundamental questions that remain to be answered conclusively include:

1. How is phase-locked neural activity transformed into a rate-place representation of pitch?
2. Where does this transformation take place, and what types of neurons perform the analysis?
3. Are there separate pitch mechanisms for resolved and unresolved harmonics?
4. How does the pitch mechanism(s) interact with the grouping mechanism(s) so that the output of one influences the processing of the other and vice versa?
5. How and where is the information about pitch used in object and pattern identification?

These questions may be answered using several techniques. Neurophysiology and brain imaging techniques may provide important clues as to mechanisms and locations. A clear demonstration of a periodotopic representation, in which the activity of different neurons/brain regions is determined by pitch *independent* of frequency content, would be a huge step forward, and there are encouraging developments in this direction (Winter, Chapter 4). Of similar importance would be the identification of a neuron that performs a synchrony-to-rate conversion with enough resolution to satisfy the psychophysicists. It may be that such neurons have already been documented, and this is where the modelers come in. We may not have a clear idea of what a pitch neuron should look like, but if we can build a model of pitch based on the known responses of particular auditory neurons that accounts for the behavioral data (including the perceptions of hearing-impaired listeners and cochlear implantees), then that will be good evidence that we are on the right track.

Recent behavioral experiments have greatly improved our understanding of

grouping mechanisms, and it is likely that they will continue to do so. Again, modelers can help illuminate the significance of the data with regard to the processing algorithms used by the auditory system. Comparisons with the physiology may also inform, as it is possible that some of these algorithms are implemented at a fairly low (and more easily probed) level in the auditory pathway. Similarly, imaging studies can probe the brain regions involved in grouping and identification.

Although it may seem obvious, it is important to emphasize that our progress in this area is dependent on collaboration between the different disciplines of psychophysics, neurophysiology, imaging, and modeling. The more avenues we can find for communication, the better our prospects will be.

## *References*

- ANSI (1994) American National Standard Acoustical Terminology. New York: American National Standards Institute.
- ASA (1960) Acoustical Terminology SI, 1-1960. New York: American Standards Association.
- Burns EM, Viemeister NF (1976) Nonspectral pitch. *J Acoust Soc Am* 60:863–869.
- Hartmann WM (1997) Signals, Sound, and Sensation. New York: Springer-Verlag.

# The Psychophysics of Pitch

CHRISTOPHER J. PLACK AND ANDREW J. OXENHAM

## 1. Introduction

Pitch is a perceptual, rather than a physical, variable. It follows that pitch processing in the auditory system can be understood only by reference to our perceptions. This chapter provides an overview of human psychophysical research on stimuli that elicit a pitch percept. The results are discussed with reference to various theoretical positions that have been taken over the years. When developing a model of pitch perception, or when identifying a cell type or brain region that may be involved in pitch perception, it is important to ensure that the results are consistent with the wide range of psychophysical observations, and not to focus on a single property of pitch that may provide an easy solution. With this in mind, the chapter emphasizes the diversity of pitch phenomena.

### *1.1 Methodology*

The aim of human psychophysical research is to improve our understanding of sensory systems by performing behavioral measurements on humans. Usually this involves tasks in which participants are required to make comparisons between sensory stimuli. It is possible to measure, for example, the smallest detectable difference along a specific physical dimension, such as frequency, or to find two stimuli that differ physically, yet are matched along some perceptual dimension, such as pitch. In addition, listeners are usually required to make discriminations or comparisons in response to brief sounds presented over headphones in an acoustically isolated environment.

The smallest detectable frequency difference between two pure tones is often referred to as the “frequency difference limen” (FDL or DLF). Similarly, the smallest detectable difference in fundamental frequency (F0) between two complex tones is sometimes called the “fundamental frequency difference limen” (F0DL). Difference limens can be measured using an adaptive procedure, in which the frequency difference between two tones is reduced as the listener makes correct responses, and increased as the listener makes incorrect responses.

The frequency differences at the “turnpoints” between decreasing and increasing frequency differences can be averaged to find the frequency difference at which the listener produces a predetermined level of performance in terms of percentage of correct responses. Results can be plotted in terms of the absolute frequency difference (in Hertz), or in terms of a relative frequency difference (the FDL as a proportion or as a percentage of the baseline frequency).

Alternatively, the frequency difference between two tones can be fixed for a number of trials, and the percentage of correct discriminations recorded. This is called the “method of constant stimuli.” It is thought that frequency discrimination is limited by variability, or noise, in the representation of frequency in the auditory system. One way to characterize this variability is to record the percent correct responses for a number of frequency differences and plot a “psychometric function” of correct responses against frequency difference. The greater the internal variability, the shallower this function will be. The percent correct scores can be converted into the discrimination index,  $d'$ , which is a measure of the difference between the means of the internal representations of the tones (i.e., some measure of physiological activity), divided by the standard deviation of the probability distributions of these representations (Green and Swets 1966). The measure  $d'$  is very useful when investigating how information is combined in the auditory system—how performance improves with duration, for example, or how simultaneous information is combined across different frequency regions.

In pitch research, listeners are often required to make comparisons between stimuli that differ along various dimensions. These techniques can be used to estimate the effects of stimulus manipulations on the pitch that is heard. For example, listeners may be required to vary the frequency of a pure tone until it matches the pitch of another pure tone with a different level, or to vary the F0 of a complex tone until it matches the pitch of another complex tone with a mistuned harmonic. Listeners may also be asked to make judgments of musical intervals, for example, by adjusting the frequency or F0 of one tone until it sounds a fifth or an octave above another tone.

## 2. Pure Tones

A pure tone has a sinusoidal variation in pressure over time. Pure tones can be regarded as the fundamental building blocks of sounds. Fourier’s theorem states that any complex waveform can be produced by summing pure tones of different amplitudes, frequencies, and phases. This insight is crucial to our understanding of the function of the peripheral auditory system, which separates out (to a limited extent) the different Fourier components of a complex sound.

Uniquely among periodic sounds, the repetition rate of a pure tone is identical to its spectral frequency. The frequency of the pure tone also corresponds to the pitch we hear, with reference to, say, the repetition rate of a complex tone. From our knowledge of the physiology of the peripheral auditory system, it is immediately apparent that there are two ways in which the frequency of a pure



tone might be represented: in terms of the pattern of excitation on the basilar membrane (e.g., the place of maximum excitation) and in terms of the temporal pattern of phase-locked firings in the auditory nerve. These two hypotheses are evaluated at the end of this section.

### *2.1 Parametric Effects on the Pitch of Pure Tones*

By definition, pitch varies with pure-tone frequency, although it has been suggested that the variation is not linear, in the sense that a given change in frequency may not produce the same change in the magnitude of the pitch sensation. The “mel scale” was derived by Stevens et al. (1937) by requiring listeners to adjust the frequency of a comparison tone until the pitch sounded half that of a standard. A frequency of 1000 Hz was used as the arbitrary reference, and a tone with this frequency assigned a pitch of 1000 mels. The scale of Stevens et al. shows that as frequency increases above 1000 Hz the mel value becomes less than the frequency value, so that a frequency of 5000 Hz produces a pitch of around 3500 mels, for example. However, a replication of this experiment by Siegel (1965) resulted in a much closer relationship between frequency and mels; in Siegel’s results half pitch was very close to half frequency. There are also theoretic reasons to doubt the validity of the mel scale. Most musicians are able to categorize musical intervals correctly over a wide frequency range. These intervals (e.g., fifths, octaves, etc.) are defined in terms of a frequency ratio. For instance, an octave is always a doubling in frequency. Given that very few musicians would claim that one octave sounds larger or smaller than another, the relationship between the nonlinear mel scale of Stevens et al. and our perception of musical pitch is tenuous at best (Houtsma 1995).

Although definitions vary (see Plack and Oxenham, Chapter 1), pitch is usually regarded as having some relationship to musical melody. If a sound does not produce a sensation of pitch it cannot be used to produce a musical melody, and it can be argued that if a sound cannot be used to produce a musical melody then it cannot be regarded as having a “true” pitch. With regard to pure tones, several studies have indicated that frequencies above about 4000 to 5000 Hz cannot be used to produce recognizable musical intervals (Ward 1954) or recognizable melodies (Attneave and Olson 1971). For those with suitable sound-production equipment, these findings can be confirmed by a casual listening test. It can surely be no coincidence that the highest note on an orchestral instrument (the piccolo) is around 4500 Hz.

Matching experiments between pure tones of different levels have revealed a limited effect of level on pitch. Below around 2000 Hz, the pitch of pure tones tends to *decrease* with increasing level. Above 2000 Hz, pitch tends to *increase* with increasing level. The maximum reported shifts are on the order of 5% to 10% (Stevens 1935), although usually the shifts are closer to 1% to 2% (Verschuure and van Meeteren 1975). There is a great deal of individual variability in these effects. Rossing and Houtsma (1986) report that for short tone bursts (40-ms duration) level increases always seem to lower the pitch, regardless of

frequency. Thus, the results suggest a possible interaction between duration and level.

The pitch of a pure tone can also be influenced by the presence of other spectral components. For example, a bandpass noise presented in the frequency region below a test tone may cause the pitch of the tone to increase (Terhardt and Fastl 1971). The effect increases with the intensity of the noise, up to a maximum of around 4%. In addition, the pitch of a mistuned partial in a complex tone is shifted slightly further upward or downward than would be predicted on the basis of the mistuning alone (Hartmann and Doty 1996; see Section 3.3.1). The pitch of the mistuned partial seems to be affected by the presence of the other components, as if the pitch were “pushed away” from the harmonic frequency (de Cheveigné 1999).

## 2.2 *Parametric Effects on the Frequency Difference Limen*

The FDL for pure tones varies in a complex way with frequency, duration, and level. For a given level and duration, the FDL in Hertz generally increases with frequency. Combining the results of several studies using long-duration pure tones at moderate levels, Wier et al. (1977) estimated that the *logarithm* of the FDL (in Hertz) is linearly related to the *square root* of frequency. Alternatively, when expressed as a *proportion* of frequency, the relative FDL decreases with frequency up to around 500 to 2000 Hz, then increases, with performance deteriorating dramatically for frequencies above around 4000 Hz (Moore 1973). Moore’s data are plotted in Figure 2.1.

Moore’s results also show that there is a strong effect of stimulus duration on the FDL. This effect is dependent on frequency, such that the change with duration (i.e., the proportional reduction in the FDL with increasing duration) *decreases* with increasing frequency up to around 4000 Hz. Importantly, there is a noticeable *increase* in the duration effect at even higher frequencies. It would make a tidy story if the relative FDL were determined by the number of periods of the pure-tone stimulus, such that a constant number of periods produced a constant relative FDL, regardless of frequency. For a limited range of frequencies (500 to 2000 Hz) and durations (6.25 to 50 ms), Moore’s data suggest that this may indeed be the case. However, the relationship certainly does not hold over the entire frequency range, breaking down badly at very low and high frequencies.

Finally, the FDL varies with sound level. At low sensation levels (close to absolute threshold) the FDL is greater than it is at moderate to high levels. The variation in performance with level (when expressed as a proportional change in the FDL) is greater for low frequencies than for high. For example, as they increased sensation level from 10 to 40 dB, Wier et al. (1977) found a decrease in the FDL from 4.3% to 0.5% at 200 Hz, and from 1.5% to 0.9% at 8000 Hz. Random variations in level between tones in a discrimination task (so that listeners have to ignore changes in gross excitation level when performing the task) have little effect on the FDL for frequencies below 4000 Hz (beyond that

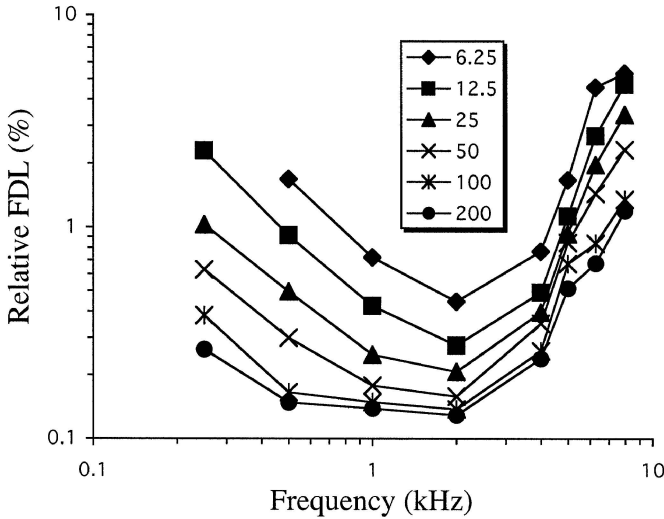


FIGURE 2.1. Pure tone frequency discrimination as a function of frequency and duration. Results are expressed in terms of the relative FDL in % ( $100 \times \Delta f/f$ ). The legend shows stimulus duration in milliseconds. Data are from Moore (1973).

predicted by the variation in pitch with level), but have a larger effect on the FDL for higher frequencies (Henning 1966; Emmerich et al. 1989; Moore and Glasberg 1989).

### 2.3 *Place versus Temporal Coding*

As suggested above, there are two obvious ways in which the pitch of a pure tone could be represented in the auditory system. First, it could be determined by the *place* on the basilar membrane that is maximally excited by the tone, or more generally by the *pattern* of excitation on the basilar membrane. This is sometimes called a “rate–place” representation since, in terms of neural activity, pitch is represented by the rate of firing of neurons responding to excitation at different places along the basilar membrane. Second, pitch could be determined by a purely temporal code, based on the property of neurons to fire in synchrony with the phase of the acoustic waveform. In effect, a pure tone of a given frequency will tend to produce action potentials separated by integer multiples of the period of the tone. A third possibility was suggested by Loeb et al. (1983). The response of the whole basilar membrane to a pure tone takes the form of a traveling wave: at a given time, different places on the basilar membrane are at different phases in their cycles of vibration. The relative phases of two given points along the traveling wave at a given time depend on the frequency of the tone. Hence frequency could, in principle, be represented by an array of coincidence detectors, with each detector responding to synchronous

activity at two specific places on the basilar membrane. A similar mechanism was suggested by Shamma (1985a,b). As pointed out by de Cheveigné (Chapter 6), this can be regarded as a version of autocorrelation (see Section 4.1.1), with the phase dispersion along the basilar membrane acting in place of a neural delay line. This mechanism also requires that neural activity is phase locked to the pattern of vibration on the basilar membrane.

Auditory-nerve recordings indicate that the ability of a fiber to phase lock to a pure tone breaks down around 5000 Hz in the cat (Johnson 1980), although this value is lower (3500 Hz or so) in the guinea pig (Palmer and Russell 1986). It has been assumed that above this frequency neurons can no longer represent the periodicity of the stimulus in terms of synchrony of firing. Some of the psychophysical results presented in this section also suggest that there may be a change in frequency coding around 4000 to 5000 Hz. First, frequency discrimination seems to deteriorate dramatically as frequency is increased above 4000 Hz (Moore 1973). Second, the effect of tone duration on the FDL increases as the frequency is raised above 4000 Hz (Moore 1973). Third, random variations in level, which might be expected to disrupt a place representation, have a substantial effect on the FDL only for frequencies of 4000 Hz and above (Henning 1966; Emmerich et al. 1989; Moore and Glasberg 1989). Finally, our perception of musical melody and our ability to recognize musical intervals breaks down above 4000 to 5000 Hz (Ward 1954; Attneave and Olson 1971).

A possible interpretation of these findings is that the frequency of pure tones may be represented in terms of phase locking (a temporal representation) for frequencies below about 5000 Hz, and purely spectrally (a place representation) for higher frequencies. In addition to the qualitative evidence, there is also quantitative support for the use of phase locking at low frequencies. Interpreting his data in terms of Zwicker's (1970) place model of frequency modulation detection, Moore (1973) argued that the FDL does not vary with frequency in the way predicted by the model. Furthermore, FDLs for short-duration tones below 5000 Hz were lower than predicted by the model, and above 5000 Hz were higher than predicted by the model (taking into account the spectral spread produced by gating the tone, and assuming that detection threshold corresponds to at least a 1-dB change in excitation). Moore's analysis suggests that changes in the excitation pattern are too small to account for frequency discrimination at low frequencies, and that another mechanism must be involved.

There are some notes of caution, however. The finding that pitch is dependent on level appears to contradict a purely temporal account, although if the temporal representation is converted into a rate representation at some stage in the auditory system then it is conceivable that the overall rate of firing in the auditory nerve might have some effect (see Moore 2003). It must be noted that the magnitude of the pitch shift is much less than would be predicted by the shift with level of the peak of excitation on the basilar membrane, which can be half an octave for high-frequency tones (McFadden 1986; Ruggero et al. 1997). At first sight, the pitch shifts produced by the presence of other spectral components (Terhardt and Fastl 1971; Hartmann and Doty 1996) do not sit happily with a

temporal account. However, de Cheveigné (1999) has shown that a time-domain model can account for the effects of other components on the pitch of a mistuned partial in a complex tone.

In summary, there seems to be a reasonable consensus at the time of writing that the representation of pure-tone frequency relies on phase locking at low frequencies, although the frequency at which the transition to a purely spatial representation occurs is a matter for debate. It has been argued recently that there is sufficient temporal information in the auditory nerve to contribute to human frequency discrimination up to frequencies as high as 10 kHz (Heinz et al. 2001a,b). If the “traditional” value of 5000 Hz is taken as the transition point, then the observation that melody recognition seems to break down for frequencies above 5000 Hz suggests that musical pitch may depend on phase locking, for pure tones at least. It could be argued that a peak or feature in the excitation pattern may not be sufficient to produce a clear musical pitch, although it is also possible that place and temporal information are combined in some way, even at low frequencies.

### 3. Complex Tones

A complex tone can be defined as any sound with more than one frequency component that evokes a sensation of pitch. However, it is possible to make a distinction between periodic (or harmonic) complex tones, and aperiodic (or inharmonic) complex tones. The former consist of a series of harmonics with frequencies at integer multiples of  $F_0$ ; the latter consist of partials that are mistuned from harmonic relationships (Hartmann 1997, p. 117). Most tonal sounds in the environment, such as vowel sounds and the sounds produced by tonal musical instruments, are harmonic complex tones, and these stimuli have been the focus of the majority of the research endeavor in pitch perception.

#### *3.1 The Missing Fundamental*

Ohm (1843) believed that the pitch of complex tones was derived from the frequency of the lowest harmonic. For almost every complex tone encountered in the environment this explanation works quite well, since the repetition rate of a complex tone is equal to the frequency of the first harmonic (or fundamental component) which is usually present in the spectra of natural sounds. Although Seebeck (1841) showed that sounds with very little energy at  $F_0$  still produced a strong pitch corresponding to the fundamental, the fact that Helmholtz (1863) favored Ohm’s explanation settled the matter for nearly a century. However, Schouten (1938) showed that removing the fundamental component completely from the acoustic stimulus did not alter the pitch and Licklider (1956) laid the matter to rest by showing that the same pitch was heard even when the frequency region that would normally be occupied by the fundamental was masked by noise. It follows that it must be possible to derive the pitch of the fundamental

from information in the higher harmonics. In the literature, this pitch has been described using many different terms, including low pitch, residue pitch, and periodicity pitch. In this chapter we refer to it primarily as periodicity pitch.

### 3.1.1 Combination Tones

Licklider set an important example for future researchers in his use of masking noise. It is now known that the cochlea's response to sound is extremely non-linear, exhibiting as much as 5:1 compression or more in high-frequency (basal) regions (Yates et al. 1990; Oxenham and Plack 1997; Ruggero et al. 1997). The nonlinearity produces intermodulation distortion products when two or more sinusoidal components are presented simultaneously (as in the case of a complex tone consisting of a number of harmonics). These "combination tones" propagate from the place of generation on the basilar membrane to the places tuned to the frequencies of the combination tones. The frequencies of the distortion products commonly observed in otoacoustic emissions (Kim et al. 1980) and in the response of the basilar membrane (Robles et al. 1997), are given by  $f_2 - f_1$  and by  $f_1 - k(f_2 - f_1)$ , where  $f_1$  and  $f_2$  are the frequencies of the physically presented components, and  $k$  is an integer. For a complex tone that has harmonic components, these distortion products are at *harmonic frequencies* (including the F0 component). It follows that even if lower harmonics are removed from the physical stimulus, they can be *reintroduced* by distortion in the cochlea (Pressnitzer and Patterson 2001). It is desirable, therefore, that in psychophysical and physiological experiments restricted to higher harmonics, a masking noise, or perhaps some other procedure, is used to render the combination tones inaudible. Researchers who do not take this precaution are open to the criticism that a listener's performance (or the response of a neuron in physiological studies) was based on combination tones rather than on the intended stimulus.

## 3.2 Dominance Region

Given that the F0 component does not have to be present in order for a pitch at the fundamental to be heard, the question then follows as to which harmonics are most important for pitch perception. Much of the research in this area has been couched in terms of defining the "dominance region" for pitch perception.

Some early work on which harmonics were of most importance involved separating the spectrum into low- and high-frequency harmonics, and discovering which group dominated the pitch percept (Plomp 1967; Ritsma 1967). Plomp (1967) presented listeners with two complex tones, in random order. One was a harmonic complex consisting of the first 12 components of a harmonic series with an F0 of  $f$ . The other was a compound complex, also consisting of 12 tones, but the lower components were harmonics of  $0.9f$  while the upper components were harmonics of  $1.1f$ . The F0 and the crossover point between the lower and upper harmonics were the experimental parameters. The reasoning was that if the lower harmonics dominated the pitch percept then the har-

monic complex would sound higher than the compound complex. On the other hand, if the higher harmonics were dominant, then the harmonic complex would sound lower than the compound complex. Plomp found that for F0s up to about 1400 Hz, the pitch was determined by the second and higher harmonics; above 1400 Hz the fundamental itself determined the pitch. For F0s up to about 700 Hz the third and higher harmonics dominated pitch judgments, while for F0s up to about 350 Hz, the fourth and higher harmonics were dominant. In no cases tested by Plomp were the fifth and higher harmonics dominant. The results, based on judgments from 14 listeners, suggest a complex interaction between F0 and spectral region: the transition point between low- and high-frequency dominance is not constant in terms of either harmonic number or absolute frequency. In very broad terms, the dominant pitch region could be viewed as incorporating the second, third, and/or fourth harmonics, except at the highest F0s, with a trend for the harmonic number at the transition to decrease with increasing F0.

Ritsma (1967), using somewhat different techniques, tested a smaller range of F0s (100, 200, and 400 Hz) and only four listeners. By using a narrower range of harmonics, he concluded that the frequency band containing the third, fourth, and fifth harmonics tended to dominate the pitch percept. However, even in the smaller range of F0s he tested, an interaction with F0 was also apparent. For instance, with a 100-Hz F0, the dominant region began between the third and fourth harmonics, whereas it tended to start at the second harmonic with a 400-Hz F0.

Both Plomp and Ritsma found that relative level did not play a large role in pitch dominance. In fact, Ritsma (1967) found that the relative contributions of components were essentially independent of level for sensation levels up to at least 50 dB, so long as the components were at least 10 dB above their absolute threshold.

Later studies attempted to narrow down the region of dominance by looking at the influence of individual components on the overall pitch of a complex. Moore et al. (1985) systematically varied the frequency of one component in a 10- or 12-component complex that was otherwise harmonic, and asked listeners to match the pitch of the complex to that of a truly harmonic complex with the same number of components. They found that individual mistuned harmonics could alter the pitch of the overall complex by a small amount and that, for shifts up to 3%, the change of the overall pitch was linearly related to the change in the frequency of the individual mistuned harmonic. On the question of which harmonics had the most influence on the overall pitch, the results were rather variable. However, some general trends emerged: for F0s of 100, 200, and 400 Hz, the most dominant harmonics tended to be the second, third, or fourth, although in some individual cases the fundamental itself was dominant; shifts in harmonics above the sixth had no measurable effect on the overall pitch. The most recent study to address this issue used a method of correlational analysis (Dai 2000). Here, listeners were presented with two successive complexes, which were nominally harmonic and had the same F0. However, the frequencies

of all the harmonics were randomly varied (or “jittered”) from interval to interval with a standard deviation of 2% of the nominal frequency. On each trial listeners were asked to judge which of the two complexes had the higher pitch. By correlating the individual frequencies with listeners’ responses on a trial-by-trial basis, it was possible to derive the perceptual “weight” that listeners placed on each harmonic in making their judgments (e.g., Berg 1989; Richards and Zhu 1994). With F0s from 100 to 800 Hz, Dai (2000) found that his data were best described in terms of a dominant frequency region, rather than dominant harmonic numbers. Specifically, he found that harmonics closest to 600 Hz tended to dominate; for F0s of 600 Hz and above, the fundamental itself carried the most weight. No harmonics above 2400 Hz were given significant weight, a finding that is broadly consistent with Plomp’s (1967) conclusion that for F0s above about 1400 Hz, the fundamental dominated the percept. A striking difference between Dai’s (2000) results and those of Moore et al. (1985) was that his weighting functions at the lowest F0s seemed to be more narrowly tuned. For instance, at F0s of 100 and 200 Hz, Dai’s mean data show distinct weighting peaks at the sixth and third harmonic, respectively, while the mean data from Moore et al. (1985) show no single peaks, but rather dominant bands spanning at least four harmonics (see Fig. 2.2). It is not clear what accounts for these differences. Two suggestions were offered by Dai (2000). The first is that in his case listeners may have been less likely to fuse the somewhat inharmonic stimulus and so may have been more likely to respond to individual harmonics, thereby exaggerating the influence of the most salient harmonic. The second is that in the case of Moore et al. (1985), as only one harmonic was mistuned at a time, listeners’ attention may have been drawn to that harmonic, thereby artificially increasing its influence on the overall pitch, and hence broadening the apparent dominance region.

In summary, while there are substantial individual differences and differences across studies, there is broad agreement that the dominant harmonics are generally between the first and fifth and that there is a tendency for the dominant harmonic number to decrease with increasing F0 (see also Patterson and Wightman 1976). There is evidence that for very low F0s (e.g., 50 Hz), harmonics higher than the fifth may be dominant (Moore and Glasberg 1988).

### *3.3 Synthetic and Analytic Listening: Global Pitch and the Pitch of Individual Harmonics*

When presented with a complex tone, such as a note on the piano or clarinet, we generally hear a single sound, with a “global” pitch corresponding to the F0. However, under the right circumstances, we are able to “hear out” individual partials from within a harmonic tone complex. As discussed in more detail later, the first five to ten harmonics can be heard in this way, depending on the F0 and the method used to measure the threshold. Listening to the global pitch and listening to the pitch of the individual harmonics have been termed synthetic and analytic listening, respectively. This section deals with the pitch of har-



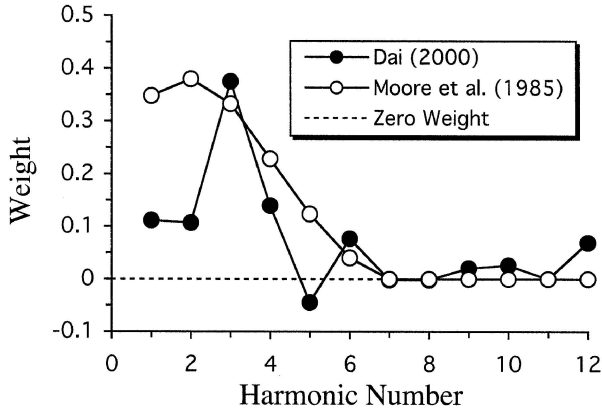


FIGURE 2.2. The results of Dai (2000) and of Moore et al. (1985) showing the relative contribution of an individual harmonic to the pitch of a complex tone as a function of harmonic number. The F0 was 200 Hz.

monics embedded within a complex, and with how these pitches may contribute to the overall pitch of the complex.

### 3.3.1 Pitch Matches to Individual Harmonics

How the pitches of individual components are perceived is potentially important for theories and models of pitch. For instance, Terhardt's (1974) model specifically assumes that it is the perceived pitches of the individual components that are combined to form the global pitch. There has been some debate as to whether the pitch of a component within a complex is the same as the pitch of a corresponding pure tone presented in isolation. Terhardt (1971) reported that this was not the case; he found that the pitch of the fundamental component was shifted downward somewhat and the pitches of the 2nd to the 4th harmonics were shifted upward by as much as 3% or 4%. Terhardt explained these shifts in terms of the excitation patterns produced by the complexes: mutual masking between neighboring harmonics alters the shape of the peak produced by the individual harmonics, leading to corresponding shifts in pitch. Within this framework, the difference in shift between the fundamental and the upper harmonics is attributable to the somewhat asymmetric nature of excitation patterns (for the upper harmonics) and to the fact that the fundamental is masked only from above (i.e., from its higher neighbors). However, later studies failed to replicate these pitch shifts, and generally found that the pitches of components within a complex are perceived as the same as when the components are presented in isolation (Peters et al. 1983; Hartmann and Doty 1996). These failures to replicate cast some doubt on the spectral explanation of the pitch shifts and on the pitch shifts themselves as an explanation for other pitch-related phenomena (Terhardt 1974, 1979; Terhardt et al. 1982a,b).

Hartmann and colleagues (Hartmann et al. 1990; Hartmann and Doty 1996; Lin and Hartmann 1998) investigated the pitches of harmonics that are mistuned from their nominal frequencies. They found an interesting pattern of results, whereby the *pitch* of the harmonic was shifted more than the *frequency* of the harmonic. In other words, if the mistuning of a harmonic was negative, the pitch was matched to a frequency lower than that of the mistuned harmonic; if the mistuning was positive, the pitch was matched to a frequency higher than that of the mistuned component. The magnitude of the pitch shift was 1% to 2%. Their results are not consistent with a place or excitation-pattern model of pitch shifts (Terhardt et al. 1982b), which predicts a positive pitch shift regardless of whether the mistuning is negative or positive (Hartmann and Doty 1996). To explain their results, Hartmann and Doty initially used a model based on interspike intervals (ISIs) in auditory-nerve fibers tuned to frequencies close to the mistuned harmonic. The underlying idea was that the pattern of ISIs would be influenced not only by the component itself, but also by neighboring components. For instance, if the harmonic was subjected to a positive mistuning, auditory-nerve fibers responding best to it would be more influenced by its upper neighbor than its lower neighbor, leading to an increase in estimated frequency. Although this scheme produced a reasonable account of the effect, its validity was placed in doubt by the later finding of Lin and Hartmann (1998) that the same pattern of mistuning was found even when harmonics neighboring the mistuned component were omitted from the stimulus. They concluded that, although the local spectrum around the mistuned harmonic played some role, the dominant effect relied on more global processes. In particular, they described their results in terms of a harmonic template, which would act to enhance the contrast between components that did and did not match the template for a given F0. In other words, if a component did not quite match one of the expected harmonic frequencies, the perceptual distance (or pitch difference) between it and the expected frequency would be increased. Studies that have modeled aspects of the pitch of mistuned harmonics are described in a later chapter (de Cheveigné, Chapter 6).

### 3.3.2 Fundamental Frequency Discrimination: Global or Local Comparisons?

How do we tell when the F0 of a complex has changed? According to temporal models of pitch perception (e.g., Meddis and O'Mard 1997; see de Cheveigné, Chapter 6), the stimulus periodicity information is pooled across all frequencies to produce a single estimate, which can then be compared with that from another stimulus. According to place-based or pattern-recognition models, the frequencies of the individual harmonics are estimated and are used to calculate the global pitch. In this case, there are at least two theoretically distinct methods by which a comparison of F0s could be made. Either the global estimates of F0 could be compared across the two stimuli or, if the same harmonics are present in both, the frequencies of the harmonics could be compared on an individual basis and the information from these multiple comparisons could be

combined. Using an optimum processor model, Goldstein (1973) tested the idea that F0 discrimination for harmonic complexes could be explained by an optimal combination of the information from each harmonic. He found that F0DLs for complex tones were greater than predicted by the FDLs of their constituent harmonics and concluded that F0 discrimination must also involve a more central internal noise source. Moore et al. (1984) reexamined Goldstein's idea, but suggested that a comparison of F0DLs with pure-tone FDLs in quiet might be inappropriate. Instead, they measured FDLs for individual harmonics embedded within the rest of the harmonic complex. They found that the presence of the other harmonics made performance substantially worse and that when these pure-tone FDLs were used to predict F0DLs for the overall complex, it was no longer necessary to postulate an additional internal noise within the framework of the optimum processor model.

Faulkner (1985) interpreted the findings of Moore et al. (1984) differently. He argued that a true F0 discrimination task would need to rule out the possibility that listeners were simply making frequency comparisons of the individual harmonics, without comparing the global (or periodicity) pitch. Faulkner's experiments showed that F0DLs were considerably worse when two complexes had no harmonics in common than in the more usual case of having the same harmonics present in both complexes. He concluded that "true" F0 discrimination was considerably worse than predicted by individual pure-tone FDLs, and that more traditional experiments, using the same harmonics in the two complexes, were measuring listeners' abilities to discriminate the individual component frequencies rather than the global pitch.

This conclusion is somewhat counterintuitive, given that listeners almost invariably report hearing the F0, rather than a collection of individual harmonics, when presented with a harmonic complex tone. On the other hand, introspection can often be misleading and cannot be used as strong evidence in favor of one position or another. Substantial light was shed on the issue by Moore and Glasberg (1990). Their experiments provide quite strong empirical support for the notion that listeners are using the F0 itself, rather than simply the individual harmonic frequencies, when performing F0 discrimination, even when the same harmonics are present in both complexes. First, they demonstrated that even when two harmonic complexes shared the first six (and most dominant) harmonics, a deterioration in performance resulted from the complexes having different higher harmonics (one had harmonics 7, 9, and 12 while the other had 8, 10, and 11). Second, they showed that listeners could not ignore the F0, even if it was advantageous to do so. The experiment involved two complexes in which only the frequency of the lowest component of each was varied. In one condition, the higher components were the same for the two complexes; in the other the higher components were harmonics from different F0s, with the lowest component being common to both F0s. Performance was much worse in the condition with different F0s. Finally, Moore and Glasberg showed that a comparison of multiple frequencies that were not harmonically related led to worse performance than when the frequencies were harmonically related. The results

clearly showed that the global pitch elicited by the F0 had a significant effect on performance; in the second example it interfered with performance and in the third example it aided performance. In the first example the detrimental effect of different higher harmonics, which themselves have very little effect on the overall pitch, suggests that the deterioration in performance found in F0 discrimination tasks when no harmonics are in common may be better ascribed to a “distraction” effect produced by differences in timbre, rather than an inherent noise associated with comparing complex tones with different F0s.

More recent work by Hafter and Saberi (2001) on the effects of cue tones on signal detection also suggests a perceptual role for the pitch of the fundamental over and above that produced by the spectral similarity of the harmonics. They showed that a harmonic three-tone target, with a random F0 and a random selection of harmonics, was more easily detectable in a noise background than an inharmonic random-frequency three-tone target. They then proceeded to investigate the effect of informing subjects of the target frequencies by using suprathreshold cue tones. They found that presenting the cue tones at the frequencies of both the inharmonic and harmonic three-tone targets improved detection. However, they also showed that presenting cues at different but harmonically related frequencies to those of the harmonic targets improved performance also. Finally, the effects of spectral and harmonic similarity were found to be additive, such that cue tones that were spectrally identical to the harmonic targets produced the highest level of performance. The level of performance was similar to that predicted by a simple detection-theoretic model in which F0 and spectral cues were considered independent sources of information. The results from both the cued and uncued conditions suggest that the global pitch provides a level of analysis, or representation, that is different from (and possibly orthogonal to) that provided by the individual spectral components.

### 3.4 *Resolved and Unresolved Harmonics*

#### 3.4.1 Defining Resolvability

It is known that the absolute bandwidth of the auditory filters increases with center frequency. Glasberg and Moore (1990) estimated that the equivalent rectangular bandwidth (ERB) of the auditory filter (in Hertz) is given by:

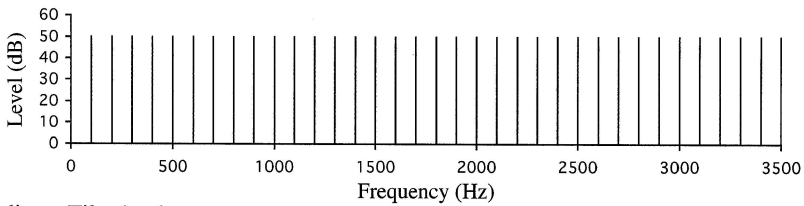
$$\text{ERB} = 24.7 (0.00437f_c + 1) \quad (1.1)$$

where  $f_c$  is the center frequency of the filter (in Hertz). For high frequencies the ERB is approximately proportional to center frequency. At 1000 Hz the ERB is about 130 Hz, that is, around 13% of the center frequency.

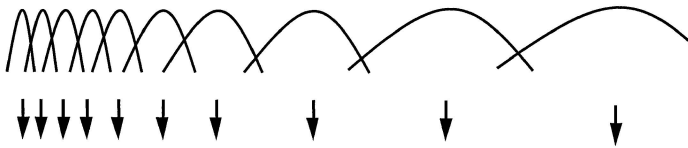
Whereas the auditory filters become broader with frequency, the component spacing in a complex tone is usually constant (and equal to F0). It follows that the spacing between harmonics, in units of auditory-filter bandwidths, decreases with increasing harmonic number; lower harmonics are separated out in the cochlea (i.e., they excite distinct places on the basilar membrane) and are said

to be “resolved,” whereas the higher harmonics are not separated out by the cochlea and are said to be “unresolved.” Figure 2.3 shows a simulated excitation pattern (the level of excitation on the basilar membrane as a function of center frequency) for a 100-Hz F0 complex with equal-amplitude harmonics (Glasberg and Moore 1990). It can be seen that the first few harmonics produce distinct peaks in the excitation pattern. As harmonic number is increased, the size of the peaks decreases relative to the troughs between them. For the high, unresolved harmonics, several harmonics interact at each place on the basilar membrane, and consequently there is little variation in excitation with center

### Input Spectrum



### Auditory Filterbank



### Excitation Pattern

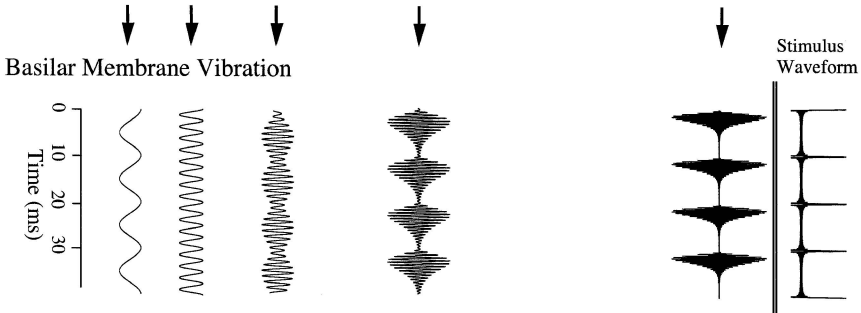
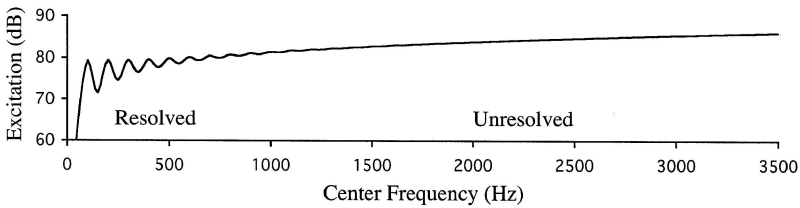


FIGURE 2.3. A schematic spectrum, excitation pattern, and simulated basilar membrane vibration for a complex tone with an F0 of 100 Hz and equal-amplitude harmonics.

frequency around each harmonic. However, whereas places on the basilar membrane responding to the lower harmonics show a sinusoidal pattern of vibration at the frequency of the harmonic, places responding to (several) higher harmonics show a complex pattern of vibration that repeats at a rate corresponding to the spacing between the harmonics (which equals  $F_0$ ).

Spectral resolvability depends more on harmonic number than on frequency per se. For example, if the repetition rate of the complex is doubled then the harmonics are spaced twice as far apart. However, each harmonic is doubled in frequency and is therefore shifted to a place on the basilar membrane where the auditory filters are approximately twice as broad. These two effects tend to cancel out, so that the resolvability of a given harmonic number does not change substantially with  $F_0$ , at least for  $F_0$ s above about 100 Hz. This relationship would be exact if the bandwidth of the auditory filters were directly proportional to center frequency.

The harmonic number at which the transition from resolved to unresolved occurs is a matter of some debate, and it depends on how resolvability is defined. Consider the excitation pattern plotted in Figure 2.3. At what harmonic number can it be said that the bump in the pattern is insufficient to constitute effective separation of the harmonic from the rest of the complex? Perhaps the most direct definition is based on perceptual separation: for a harmonic to be resolved, a trained listener must be able to “hear out” the harmonic as a pure tone with a distinct pitch. This can be measured by requiring the listener to make a frequency comparison between a pure tone and a harmonic in a complex tone. Most studies suggest that this comparison is possible for harmonics up to around number 5 to 8 (Plomp 1964; Plomp and Mimpen 1968; Moore and Ohgushi 1993), but recent results suggest that this may be possible for harmonics up to number 10 if attention is drawn to the harmonic by gating it on and off (Bernstein and Oxenham 2003).

A less direct definition was proposed by Shackleton and Carlyon (1994). When the harmonics in a complex are presented so that the positive-going zero crossings (the times at which the amplitude crosses zero between a trough and a subsequent peak in the sinusoidal waveform) of the individual harmonics are coincident, the harmonics are said to be in sine phase. The resulting waveform has an envelope that repeats at the  $F_0$ . If, however, the harmonics are alternated between sine phase and cosine phase (so that the zero crossings of the odd-numbered harmonics are aligned with the peaks of the even-numbered harmonics) the resulting envelope has a repetition rate of *twice* the  $F_0$ . This is known as alternating, or ALT, phase (see Fig. 2.4). It turns out that as the lowest harmonic number in the ALT complex is raised above number 10 or so, the periodicity pitch of the complex is an octave higher, corresponding to twice the  $F_0$ . The implication is that when the harmonics are resolved, the phase relationship between them is irrelevant since the harmonics do not interact significantly in the cochlea. However, when three or more harmonics excite the same place on the basilar membrane (i.e., are unresolved), the resulting pattern of vibration will reflect the phase relationship between them. It is suggested that

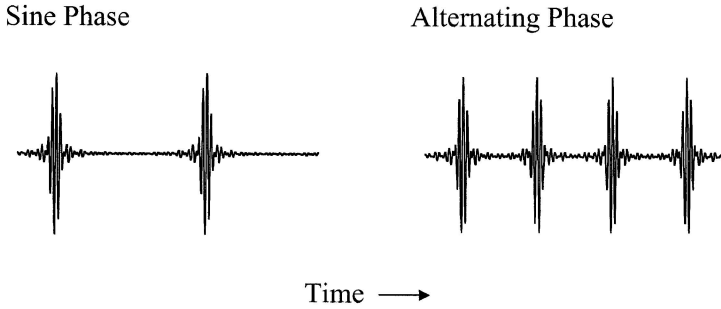


FIGURE 2.4. An illustration of a brief section of the waveforms of sine phase and alternating phase complexes, similar to those used by Shackleton and Carlyon (1994). These complexes have the same  $F_0$  (125 Hz) and the same harmonic numbers, but the pitch of the complex on the right is an octave higher than the pitch of the complex on the left. Both complexes were filtered between 3900 and 5400 Hz.

periodicity pitch is related to the repetition rate of the *temporal envelope* of these interacting harmonics, and not to  $F_0$  (see Flanagan and Guttman 1960 for earlier work manipulating the temporal envelope of harmonic complexes).

Finally, resolvability can be defined in terms of  $F_0$  discrimination. It has been observed that it is much easier to discriminate the  $F_0$ s of complexes containing low harmonics than the  $F_0$ s of complexes containing just high harmonics. Houtsma and Smurzynski (1990) measured  $F_0$  discrimination for a group of 11 successive harmonics for an  $F_0$  of 200 Hz (see Fig. 2.5). As the number of the lowest harmonic was increased from 7 to 13 there was a dramatic increase in the relative F0DL from around 0.25% to around 2.5% of  $F_0$ , with performance remaining roughly constant as the lowest harmonic number was increased above 13. It was argued that this jump in the F0DL reflects the transition from a complex containing some resolved harmonics to a complex containing no resolved harmonics. Similar experiments carried out by others using different  $F_0$ s have confirmed that the deterioration in performance is due not to the increasing absolute frequency, but to the increase in the lowest harmonic number present (Carlyon and Shackleton 1994; Shackleton and Carlyon 1994; Kaernbach and Bering 2001; Bernstein and Oxenham 2003).

These experiments suggest that, for most  $F_0$ s used experimentally, the harmonic number that marks the transition from resolved to unresolved is not less than 5 but no greater than 10. However, the transition point will depend on  $F_0$  to some extent. An inspection of Eq. (1.1) reveals that as center frequency is decreased, the ERB expressed as a proportion of center frequency *increases*. It follows that for low  $F_0$ s (below around 100 Hz) the harmonics are not as well resolved in the excitation pattern as they are for higher  $F_0$ s. The transition between resolved and unresolved will occur, therefore, at a lower harmonic number. In effect, resolvability may be defined in terms of the bandwidth of the auditory filter. For example, Moore and Ohgushi (1993) found that listeners

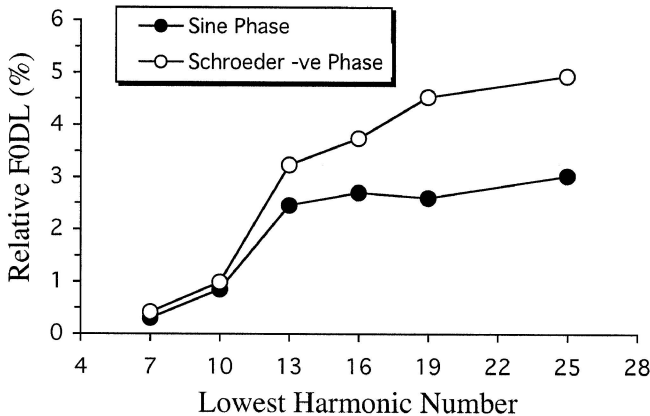


FIGURE 2.5. The results of Houtsma and Smurzynski (1990) showing the F0DL (as a percentage of F0) for a group of 11 successive harmonics with a nominal F0 of 200 Hz, as a function of the lowest harmonic number in the group. Harmonics were presented in either sine phase or in negative Schroeder phase, in which the phase relationships between harmonics were selected to produce a relatively flat envelope on the basilar membrane.

could determine whether a pure-tone probe was higher or lower than a component in an inharmonic complex at around 75% correct when the spacing between the components was 1.25 times the ERB. Similarly, Shackleton and Carlyon (1994) estimated that harmonics are resolved when there are fewer than two within the 10-dB bandwidth of the auditory filter, as defined by Glasberg and Moore (1990), and unresolved when there are more than 3.25 within the 10-dB bandwidth of the auditory filter.

From the results presented in Section 3.2 it can be seen that the region of harmonic resolvability may not coincide exactly with the region of dominance. However, it is true to say that resolved harmonics, when present, provide a greater contribution to the overall pitch than unresolved harmonics, at least for F0s of 100 Hz and above.

#### 3.4.2 Is F0 Discrimination Dependent on Resolvability or Harmonic Number?

The previous section outlined a number of different measures that seem to converge on the idea that the first 5 to 10 harmonics may be peripherally resolved. The fact that this limit coincides well with a transition between good and poor F0 discrimination suggests that good F0 discrimination requires the presence of some resolved harmonics. Though they may be necessary, the question remains whether resolved harmonics are sufficient to produce good F0 discrimination. A recent study suggests not. Bernstein and Oxenham (2003) repeated part of Houtsma and Smurzynski's (1990) study, with the addition of a "dichotic" condition, in which the odd harmonics were presented to one ear and the even



harmonics to the other. They first confirmed that the dichotic presentation doubled the number of harmonics that could be heard out individually, or resolved. As might be expected, because the frequency spacing between adjacent components in each ear was doubled, listeners were now able to hear out the first 15 to 20 harmonics of 100- and 200-Hz F0s. However, when these complexes were used to measure F0 discrimination as a function of the lowest harmonic present, performance was very similar to that found in the diotic condition, in which all components were presented to both ears (see Fig. 2.6). In other words, listeners were not able to make use of the additional resolved components to improve F0 discrimination. This shows that presenting higher components in such a way that they are also resolved does not improve performance. Similar results were found for two-component stimuli by Houtsma and Goldstein (1972; see Section 3.5.3) in normal-hearing listeners and by Arehart and Burns (1999) in hearing-impaired listeners (see Moore and Carlyon, Chapter 7).

The inability of higher harmonics to contribute to the pitch percept, even if they are peripherally resolved, has some interesting theoretical implications. From the perspective of spectral theories of pitch (de Cheveigné, Chapter 6) it suggests that harmonic templates, if they exist, are formed only of the lower harmonics, which are normally resolved. This is consistent with the idea that harmonic templates can build up through exposure to harmonic sounds (Terhardt 1974) or even to any broadband sounds (Shamma and Klein 2000). In both these cases, one requirement for such templates to emerge is that individual harmonics are normally spectrally resolved.

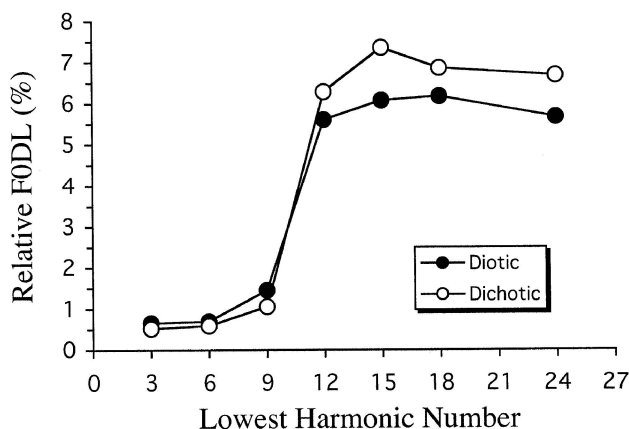


FIGURE 2.6. A “grand mean” of the results of Bernstein and Oxenham (2003) across both F0s (100 and 200 Hz) and phase relationships. The figure shows the F0DL (as a percentage of F0) for a group of 12 successive harmonics as a function of the lowest harmonic number in the group. Either all harmonics were presented to both ears (diotic) or harmonics were alternated between the left and right ears (dichotic) so that the harmonic spacing in each ear was twice the F0.

### 3.5 *Existence Regions*

#### 3.5.1 Upper Limits of Pitch

In Section 2.1 it was noted that pure tones above about 4000 to 5000 Hz do not elicit a melodic pitch. There are also limits to the ability of harmonic complex tones to carry usable pitch information, although here there are two relevant dimensions, spectral content and F0. One of the earliest studies on the so-called “existence region” of pitch used rather minimalistic stimuli, consisting of a sinusoidally amplitude-modulated tone, which has just three components (Ritsma 1962). The task was subjective: listeners adjusted the modulation depth or, equivalently, the level of the two outer components relative to that of the central component to the point at which the periodicity pitch could no longer be heard. The results varied somewhat across the three subjects tested, but the trends were rather similar. The center frequency at which the periodicity pitch could no longer be heard, even with 100% modulation depth, depended to a large extent on the F0, or (equivalently) the modulation frequency. For an F0 of 100 Hz, a pitch could be heard up to a center frequency of 2500 Hz, or the 25th harmonic; for an F0 of 200 Hz, the limit occurred around 4000 Hz, or the 20th harmonic; and for an F0 of 500 Hz, the limit occurred around 5500 Hz, or the 11th harmonic. At even higher F0s, performance rapidly deteriorated, so that at an F0 of 800 Hz, two of his three subjects reported not being able to hear the periodicity pitch even with a center frequency of 3200 Hz, or the 4th harmonic. Thus, as with the dominance region, the existence region seems to be constant in terms of neither absolute frequency nor harmonic number. However, the finding that no center frequencies above 6000 Hz produced a periodicity pitch with a three-component complex has been interpreted as evidence for a fundamental limit in the ability to hear periodicity pitch based on components with frequencies above about 6000 Hz.

#### 3.5.2 Lower Limits of Pitch

Human listeners are sensitive to periodicity over a very large range of repetition rates. At very low rates of 10 Hz or less, the individual clicks in a train of clicks are heard. At 100 Hz, the percept is one of a “buzzy” tone with a clear pitch corresponding to 100 Hz. At some point in between, therefore, lies the lower limit of pitch. The exact point of that transition depends somewhat on the operational definition of pitch. The most recent attempts to quantify the lower limits of pitch have involved both rate discrimination, or F0DLs (Krumholz et al. 2000), and melody discrimination (Pressnitzer et al. 2001). Pressnitzer et al. (2001) used a four-note random melody and asked listeners to judge which one of the four notes was altered by a semitone in a second presentation. All the notes were within a five-semitone range. For broadband harmonic complex tones (in cosine phase) the lowest range of F0s over which listeners could perform the task with around 70% accuracy was from about 32 to 40 Hz. Interestingly, the lowest note of this range (32 Hz) is rather close to the lowest

note found on most pianos (A0, 27.5 Hz). Although some organs have lower notes, these are rarely used in isolation and are generally thought to be more for musical “effect” or atmosphere than for carrying melody. As expected, based on the results of Ritsma (1962, 1963) and others, Pressnitzer et al. (2001) also found that the lower limit of pitch depended on the spectral region in which the stimuli were presented. Using a constant 600-Hz-wide band of harmonics, they found that the lower limit of pitch increased from around 35 Hz with a lower cutoff frequency of 200 Hz, to around 300 Hz with a lower cutoff frequency of 3200 Hz. Also consistent with Ritsma (1962), they found that their melody task was impossible with a lower cutoff frequency of 6400 Hz.

Krumbholz et al. (2000) measured rate (or F0) discrimination thresholds for conditions very similar to those studied by Pressnitzer et al. (2001). Although a direct comparison between melody discrimination and simple F0 discrimination is not straightforward, the patterns of results from the two tasks were reasonably similar. It is interesting to note that both studies found limits that were generally well outside the region where harmonics are considered to be spectrally resolved, so that pitch judgments were most likely mediated by temporal mechanisms. This finding is in line with those of Moore and Rosen (1979) and Kaernbach and Bering (2001). Both studies found that the pitch produced by unresolved harmonics, although weaker than that produced by resolved harmonics, was nonetheless capable of carrying information about musical intervals and melodies.

### 3.5.3 Effects of Number of Components: Dichotic and Sequential Presentation

How many components does it take to form a periodicity pitch? Much of the early work into periodicity pitches, and the limits thereof, was done using three components (see above). However, it is generally accepted that more components produce a stronger, or more salient, pitch. Even so, three components do not represent the limit for the perception of periodicity pitch. Smoorenburg (1970) showed that when two pairs of tones (1800 and 2000 Hz and 1750 and 2000 Hz) were presented sequentially, about half the 42 listeners heard the pitch go down (presumably following the spectral pitch of the lower component) while the other half of the listeners heard an upwards pitch movement, in line with the fundamental frequencies of 200 and 250 Hz, respectively.

Houtsma and Goldstein (1972) also studied the pitch produced by two-component complexes. They showed first that the pitch derived from two-component complexes was sufficiently salient to permit musical interval recognition, and second that the pitch percept remained as strong (in terms of musical-interval recognition performance) when the two components in each complex were presented to opposite ears. The second finding is especially significant in terms of understanding the mechanisms of pitch. As discussed later in the book (de Cheveigné, Chapter 6), Schouten’s (1940) proposal for a pitch mechanism involved calculating the period of a waveform comprising two or more components. As such, the theory relied on components interacting within

the cochlea, a condition that was not met in Houtsma and Goldstein's experiment, where the two components were presented to opposite ears and so did not interact peripherally at all. Thus, their results disprove Schouten's hypothesis that peripheral interaction of components is necessary for complex tone pitch perception. Another important finding of Houtsma and Goldstein (1972) was that the ability of two adjacent harmonics to convey pitch decreased with increasing harmonic number. The best performance was achieved for F0s between 200 and 300 Hz, and even there performance was poor when the lowest harmonic numbered 8 or higher. The fact that the upper limit was the same for both monaural and dichotic conditions suggests that performance was not limited by the peripheral resolvability of the components (see Section 3.4.2).

Two adjacent components are the theoretical minimum from which to derive an unambiguous periodicity pitch. However, Houtgast (1976) showed that under some circumstances, in the appropriate context, even a single upper harmonic could elicit a periodicity pitch. In his experiment, the reference interval contained a complex consisting of the harmonics 2 to 4 and 8 to 10. The other interval consisted of one, two, or three harmonics selected from harmonics 5, 6, and 7. A 3% F0 difference was always present between the two complexes and listeners had to decide whether the F0 had increased or decreased. Houtgast's results provide one example of where the addition of noise improves performance dramatically: he found that a pink noise, presented at a level such that each tone component was about 6 dB above its masked threshold, improved discrimination in all conditions. The improvement was especially dramatic when the second stimulus consisted of only one harmonic; when no noise was present, performance was near chance for most listeners, but in the presence of noise, performance improved to the extent that more than 50% of listeners scored more than 80% correct. It seems that the clear pitch in the first interval primed listeners so that they associated the single tone in the second interval with a very similar pitch. The noise may have facilitated this process by making the presence of the missing harmonics seem "plausible" to the auditory system. In other words, lacking evidence to the contrary, the ecologically most likely scenario is that the two successive complexes contain the same harmonics and the harmonics that are not perceived are simply masked by the noise.

A similarly beneficial effect of background noise was found by Hall and Peters (1981). They asked whether a periodicity pitch could be extracted from components that were presented successively, instead of simultaneously. In a paradigm similar to that used by Smoorenburg (1970) they presented short successive bursts of 600, 800, and 1000 Hz, followed after a pause by successive bursts of 720, 900, and 1080 Hz. If listeners heard primarily the spectral pitch of the components, they would tend to respond that the second interval was higher. On the other hand, if they heard the periodicity pitch (F0s of 200 and 180 Hz, respectively), they would respond that the first interval was higher. Their results were very clear: in the absence of noise, listeners responded almost exclusively to the spectral pitch. When the tones were presented in noise at 6 dB above masked threshold, listeners responded almost exclusively to the pe-

riodicity pitch. It seems that the noise may have promoted integration over time by making it plausible that the harmonics were all present throughout the interval, rather than being three separate sound events. When no noise was present, it may be that any integration of pitch information was “reset” with the onset of each new tone (see Section 6.2.2).

## 4. Unresolved Harmonics and Stochastic Stimuli

From the work on the dominant region described earlier it is reasonable to conclude that unresolved harmonics are relatively unimportant in determining pitch, at least for F0s of 100 Hz and above. In the “real world” most of the complex tones that we hear contain *resolved* harmonics. Why then do unresolved harmonics merit a large section of the chapter? One of the reasons they are worthy of attention is that the pitch of these harmonics must be derived from a temporal representation, such as the repetition rate of the waveform produced by the interaction of several harmonics on the basilar membrane (see Fig. 2.3). By definition, no place cues are available as to the frequencies of the individual harmonics. Unresolved harmonics provide a controlled way of investigating temporal processing in the auditory system. Another reason for investigating unresolved harmonics is that, because of their poorer frequency selectivity, hearing-impaired listeners (as well as cochlear-implant users) may rely more on unresolved harmonics to derive pitch (see Moore and Carlyon, Chapter 7). It is important that the mechanisms by which they do this are understood. Finally, the pitch of unresolved harmonics may be important for auditory grouping. For example, F0 differences between the first and second formants in speech can give the impression of two sound sources, even when the second formant contains only unresolved harmonics (Darwin 1992). It seems likely, therefore, that the auditory system has a real interest in deriving the pitch of unresolved components, beyond the dubious utility of performing well in psychophysical experiments.

### 4.1 Experiments with Pulse Trains

When the harmonics of a complex tone are presented so that they are all in sine phase (positive zero crossings aligned) or all in cosine phase (peaks aligned) the resulting envelope has distinct envelope peaks, or “pitch pulses.” When such a complex is filtered to contain just unresolved harmonics, the envelope of the waveform is preserved to a certain extent in terms of the pattern of vibration on the basilar membrane. (Another way of generating such a complex is by highpass- or bandpass-filtering a train of clicks; a broadband cosine-phase complex is equivalent to a regular click train.) It is often assumed that these envelope peaks are represented by synchronous activity in the auditory nerve. In other words, the timing of pitch pulses in the stimulus may be a fair reflection of the pattern of activity in the auditory nerve. Although an individual neuron

may not respond to each pitch pulse, across several neurons the individual envelope peaks may be well represented. By manipulating the timing of individual pitch pulses (thereby destroying the strict harmonic relationship of the complex) researchers have been able to test temporal models of pitch perception.

#### 4.1.1 First-Order and Higher-Order Intervals

As described in Chapter 6 (de Cheveigné), several models of pitch extraction have been based on the autocorrelation function (ACF) (Licklider 1951; Meddis and Hewitt 1991). The ACF is implemented by correlating a signal with a delayed representation of itself. At time delays equal to integer multiples of the repetition rate of a waveform, the correlation will be strong. The F0 of a complex can usually be determined by taking the inverse of the delay that produces the first large peak in the ACF of the complex. Similarly, the pitch of a complex can be estimated by taking the ACF of the simulated neural activity produced by a complex (Meddis and Hewitt 1991; Meddis and O'Mard 1997).

One of the predictions of models based on the ACF is that intervening neural spikes should not have a substantial effect on the correlation, and therefore pitch strength, produced by spikes separated by a particular delay. In other words, the ACF is not sensitive to whether the optimum delay occurred between consecutive spikes ("first-order" intervals) or between spikes separated by intervening spikes ("higher-order" intervals). The ACF is described as an "all-order" interval representation, since it represents the intervals between each pulse or spike and all its neighbors, rather than just successive spikes.

Kaernbach and Demany (1998; see also Kaernbach and Bering 2001) tested this prediction by synthesizing a regular sequence of highpass-filtered clicks. Between each successive click they inserted another click at a random time. The resulting stimulus had a strong second-order periodicity (the time interval between every other click was constant) but a random first-order periodicity. Kaernbach and Demany showed that this complex was almost indistinguishable from a random sequence of pulses with the same mean pulse rate, although the ACFs of the waveforms, which are sensitive to second-order periodicity, are very different for these two stimuli. Sequences with strong *first-order* periodicity, on the other hand, were discriminated from random sequences by the listeners.

The results seem to suggest that the auditory system does not perform the equivalent of autocorrelation in order to derive pitch, but that instead the pitch mechanism processes the first-order intervals between pulses and ignores higher-order intervals. However, Pressnitzer et al. (2002) showed that the conclusions are not so clear cut if the ACF analysis is performed on a simulation of neural activity, rather than on the waveforms themselves. Because of the combined effects of auditory filtering and hair-cell transduction in the model, the second-order stimulus used by Kaernbach and Demany produced a rather *broad* peak in correlation at the delay corresponding to the second-order interval, in contrast to the sharply defined peak that was observed in the ACF of the original wave-

form. Furthermore, the random-sequence comparison used by Kaernbach and Demany also showed a broad peak at the same delay. This was because the sum of two random variables (in this case, two consecutive first-order intervals) has a broadly peaked distribution with a mean equal to twice the mean of the distribution from which the random variables are selected (in this case, the mean period between the pulses). In other words, the modeling results of Pressnitzer et al. could explain why a second-order pulse train and a random pulse train sound similar. The overall conclusion from these studies seems to be that a model based on the ACF of the physical waveform will not work for these stimuli, but that models based on the ACF of the neural response may be able account for the experimental results.

#### 4.1.2 Mean Rate and Weighted Intervals

A simple way to estimate the repetition rate of a filtered pulse train, containing only unresolved harmonics, is to divide the duration of the tone by the number of intervals between pulses and take the inverse. The “mean rate” model of pitch (Carlyon 1996, 1997) predicts that two complexes with the same duration and the same number of pulses should have the same pitch, regardless of the temporal regularity of the pulses in the sequence. Conversely, ACF models (and indeed, any models based on the detection of “common intervals”) predict that two complexes having a preponderance of the same regular interpulse intervals will have the same pitch, regardless of the actual number of pulses in each. Carlyon (1997) tested these predictions by randomly deleting pulses from pulse trains that were bandpass filtered to contain only unresolved harmonics. When two pulse trains were manipulated to have the same mean rate but different nominal F0s (and therefore, different interval distributions), listeners had difficulty in telling them apart. Conversely, listeners could discriminate two pulse trains that had the same F0 but different mean rates. Furthermore, reducing the number of pulses while keeping the F0 constant, resulted in a reduction in pitch: a 10% reduction in the number of pulses resulted in around a 4% to 5% reduction in pitch. The results suggest that mean rate has a large effect on the pitch of unresolved harmonics. However, when only a few pulses were deleted to produce a constant mean rate, listeners could still make discriminations based on F0 (common-interval) differences.

In a further study, Carlyon et al. (2002) showed that a sequence of pulses, for which the (first-order) interpulse interval alternated between 4 and 6 ms, had a pitch that was not equal to the inverse of the individual intervals of 4, 6, or 10 ms (as predicted by common-interval models), or to the simple mean rate of 200 Hz (1000/5), but corresponded instead to an interval of around 5.7 ms. Carlyon et al. accounted for their data by modifying Carlyon’s mean-rate model. They proposed a model based on an average of first-order intervals in which the contribution of individual interpulse intervals was weighted, with shorter intervals contributing less than longer intervals.

A similar approach was taken by Plack and White (2000a). They presented

listeners with a pulse train containing 10 pulses (and therefore 9 interpulse intervals). The first four and last four intervals were fixed at 4 ms, but the center interval was varied. Although the predominant interval (8 out of 9) was always 4 ms, Plack and White found that manipulating the center interpulse interval could have a significant effect on pitch. The pitch matches obtained were inconsistent with a common-interval or ACF analysis, even when the analysis was based on simulated neural activity. The pitch matches were consistent with a mean rate model, to a certain extent: Carlyon et al. (2002) were able to produce a reasonable account of the results of Plack and White using their model based on weighted intervals.

In summary, it appears that there are some stimuli containing unresolved harmonics whose pitches are not predicted by common-interval models such as the ACF. The pitch of these stimuli may correspond to a weighted mean of the (first-order) interpulse intervals. It should be noted, however, that at least some degree of regularity seems to be necessary to produce a sensation of pitch. A totally random pulse train does not have a tonal quality.

## 4.2 *Fine Structure and Envelope*

The temporal *fine structure* of a waveform refers to the rapid variations in pressure that carry the acoustic information. The temporal *envelope* of a waveform refers to the slower, overall changes in the amplitude of these fluctuations (see Hartmann 1997). For unresolved harmonics (and certain stochastic stimuli) information about periodicity is present in both the fine structure and the envelope. In this section, a collection of experiments is described that have shed light on the relative importance of these two quantities in pitch perception.

### 4.2.1 Phase Effects for Unresolved Harmonics

Section 3.4.1 described how the pitch of a complex tone consisting of unresolved harmonics can be changed by varying the phase relationships between the individual harmonics. Specifically, harmonics in alternating sine and cosine (ALT) phase produce a pitch that is an octave higher than that of a sine-phase stimulus, and an octave higher than what one would expect from the F0 or the harmonic spacing (Shackleton and Carlyon 1994). It seems possible that the perception of pitch for unresolved harmonics is dependent on the envelope of the waveform produced on the basilar membrane by the interaction of several harmonics. The repetition rate of the fine structure of the waveform (the individual variations in amplitude that determine spectral frequency) remains equal to F0, regardless of the phase relations between harmonics.

The importance of the envelope for complexes with unresolved harmonics has been observed in other studies. Houtsma and Smurzynski (1990) found that F0 discrimination and musical interval recognition depended on the phase relationships between harmonics when they were unresolved (but not when they were resolved): A “negative Schroeder” (Schroeder 1970; Kohlrausch and Sander 1995) complex, in which the phase relationships between harmonics were selected to



produce a relatively flat envelope on the basilar membrane (minimum peakiness or crest factor), produced poorer performance than a sine-phase complex (see Fig. 2.5). It seems plausible that complexes with distinct envelope peaks on the basilar membrane produce a better-defined temporal representation in the auditory system, and therefore a more salient pitch, than those with flat envelopes.

Despite these findings, it remains the case that the temporal fine structure of unresolved harmonics with low F0s is well represented in the auditory nerve (see Winter, Chapter 4). Even though pitch seems to be affected by envelope repetition rate, it is possible that the envelope periodicity may be coded by a representation of fine structure. Schouten et al. (1962) measured the pitch of sinusoidally amplitude-modulated pure tones. These stimuli have components at  $f_c - g$ ,  $f_c$ , and  $f_c + g$ , where  $f_c$  is the carrier frequency and  $g$  is the modulation frequency. If  $f_c$  is an integer multiple of  $g$ , the waveform has a harmonic structure (e.g., 1800, 2000, and 2200 Hz), and an envelope repetition rate of  $g$  (200 Hz in this case). If  $f_c$  is increased slightly, the harmonic structure is lost (e.g., 1840, 2040, and 2240), but the envelope repetition rate remains equal to  $g$ . Schouten et al. reported that increasing or decreasing  $f_c$  produced shifts in the pitch of the waveform (compared to three-component *harmonic* complexes with the same carrier) that were consistent with the intervals between peaks in the fine structure close to (but not coincident with) the envelope peaks. At face value, the pitch of these stimuli seems to be determined by the fine structure, not the envelope. However, Moore and Moore (2003) have shown recently that the “pitch” shifts obtained by Schouten for the higher carrier frequencies may have been based on increases in the spectral “center of gravity” of the excitation pattern as the carrier frequency was increased (i.e., the match was not based on the periodicity pitch). When the spectral envelope was held constant by filtering a broadband harmonic complex with a fixed bandpass characteristic, Moore and Moore found that a shift in the individual frequencies of a group of higher, unresolved harmonics by a constant amount (i.e., maintaining the component spacing) did not produce a shift in pitch. This is consistent with a pitch mechanism for unresolved harmonics based on envelope, rather than fine structure, periodicity. However, Moore and Moore found that pitch shifts were observed for “intermediate” harmonics that they considered could be just unresolved (around the ninth harmonic). Also, Hall et al. (2003) found that for three-component unresolved complexes, there was an interaction between center frequency and the strength of envelope cues, with the envelope cues becoming increasingly important with increasing center frequency. They suggested that this was the result of the influence of fine structure at the lower center frequencies. It thus remains possible that fine structure may contribute to the pitch of some complexes with unresolved harmonics.

#### 4.2.2 Amplitude-Modulated Noise

Although the sensation produced by noise might be considered highly dissimilar perceptually to the sensation produced by a pure tone or by a harmonic complex tone, it is possible to manipulate noise in order to produce a sensation of pitch.

In effect, regularities can be introduced into the otherwise random sequences of amplitudes, and sometimes these regularities can be extracted by the auditory system. By determining the kind of regularities that produce this effect, it is possible to investigate the mechanisms of pitch perception.

If a noise stimulus is amplitude modulated, so that its envelope varies periodically but its fine structure remains random, then a weak pitch can be produced. Pollack (1969) showed that a white noise turned on and off repeatedly (“interrupted noise”) can be matched to a sinusoid with a frequency equal to the interruption rate for interruption rates up to around 2000 Hz (although this was only for one listener, and it is not certain that the comparison was made on the basis of pitch). Similarly, noise that is modulated sinusoidally (SAM noise) has a pitch corresponding to the modulation frequency. Using modulation frequencies in the range 84 to 189 Hz, Burns and Viemeister (1976, 1981) demonstrated that it is possible to produce recognizable melodies by varying the modulation frequency. In other words, the sensation produced by SAM noise seems to satisfy a fairly conservative definition of pitch. Based on musical interval recognition, Burns and Viemeister estimated that the existence region for the pitch of SAM noise extends up to around 850 to 1000 Hz.

The results show that pitch can be extracted from the temporal envelope of stimuli in the absence of fine structure regularity: the fine structure of modulated white noise is random, and the long-term spectrum is flat. On the other hand, the pitch of modulated noise is not as strong as that of a harmonic complex tone, and this could be interpreted as evidence that fine structure regularity is used by the auditory system under normal circumstances.

#### 4.2.3 Iterated Rippled Noise

If a noise is delayed by  $d$  ms then added back to the original undelayed noise, rippled noise is produced. Repeating this process a number of times results in iterated rippled noise (IRN). The spectrum of IRN contains maxima at intervals of  $1/d$  kHz, rather like the harmonics of a complex tone. Highpass filtering IRN at an appropriate frequency can effectively remove spectral cues, just as a harmonic complex tone can be filtered to contain only unresolved harmonics.

For delays between about 2 and 30 ms, both unfiltered and filtered IRN have a pitch corresponding to  $1/d$  kHz. However, there are two components to the perception of IRN: a tonal sensation and a noisy sensation. As the number of iterations increases, the tonal sensation begins to dominate the noisy sensation. Patterson et al. (1996) asked listeners to match an IRN with 1 to 16 iterations to a harmonic complex standard containing varying proportions of broadband noise. They reported that the tone/noise ratio of the matching standard increased by around 4 dB for every doubling in the number of iterations. The matches were unaffected by highpass filtering the IRN at the twelfth harmonic of the delay, eliminating resolved spectral components. Following Yost (1996) they argued that the pitch strength of IRN is related to the height of the first peak of the ACF ( $1/d$  delay).

IRN is a useful stimulus for investigating temporal pitch mechanisms (see Shofner, Chapter 3, and Winter, Chapter 4), since it is possible to vary the pattern of autocorrelation by varying the number of iterations and the gain applied to the delayed noise before it is added back to the original. The results seem to support a model based on the ACF of neural activity phase locked to the *fine structure* of the waveform. However, the temporal regularity of IRN is also present in the *envelope* of the waveform. Yost et al. (1998) attempted to distinguish between these two cues. They asked listeners to discriminate an IRN stimulus in which the gain is 1 (delay-add) from one in which the gain is  $-1$  (delay-subtract). The two stimuli have identical temporal envelopes, but they differ in their temporal fine structure and spectral composition: the delay-add stimulus has spectral peaks at harmonic frequencies of the repetition rate, whereas the delay-subtract stimulus has spectral peaks located halfway between the harmonic frequencies (as for the odd harmonics of a stimulus with half the repetition rate). Yost et al. showed that these two stimuli were easily distinguishable by listeners for frequency regions up to 4000 to 6000 Hz, and even up to 8000 to 10,000 Hz in some cases. They argued that this discrimination was based on fine-structure cues, as revealed by differences between the ACFs of the two stimuli. Unfortunately, this interpretation is not absolutely clear cut, as there is a possibility that the listeners in Yost et al.'s (1998) study were in fact using spectral peaks (rather than temporal fine structure) to make their judgments. Although the stimuli were filtered to contain only unresolved spectral peaks, nonlinearities in the ear may have generated resolved spectral peaks at lower frequencies. A control experiment by Yost et al. (1998) showed that listeners were no longer able to distinguish delay-add from delay-subtract IRN when a lowpass noise was added at a spectrum level 11 to 17 dB (depending on listener) below the spectrum level of the actual stimulus. Although not interpreted in this way by the authors, these results may indicate that listeners' abilities to perform the task was obliterated once potential distortion products were masked. To our knowledge there are currently no studies addressing combination tones in IRN, and it is possible that they are less salient than for harmonic tone complexes. On the other hand, if combination tones in IRN are as salient as they are for harmonic complexes, then they may have affected performance on this task.

#### 4.2.4 Transposing Fine Structure into Envelope

At low frequencies, the auditory-nerve response to a narrow-band stimulus resembles a half-wave rectified version of the physical waveform. At higher frequencies—above the upper limit of phase locking to fine structure—the auditory-nerve response resembles only the temporal envelope (see Winter, Chapter 4). Van de Par and Kohlrausch (1997) proposed a way of transposing the fine structure of a low-frequency stimulus into the envelope of a stimulus with a high-frequency carrier. In this way, the temporal information that would normally only be available to low-frequency auditory-nerve fibers is presented

to fibers with high characteristic frequencies (CFs). In studies of binaural processing, researchers have found that subjects can extract the temporal information from the envelope of a transposed stimulus with the same accuracy as that from the fine structure of the original low-frequency stimulus, at least for frequencies up to 150 Hz (van de Par and Kohlrausch 1997; Bernstein and Trahiotis 2002). If the temporal information conveyed by transposed stimuli can be evaluated for binaural information, can it be used for pitch? Temporal models of pitch that disregard the tonotopic location of the temporal information (Cariani and Delgutte 1996; Meddis and O'Mard 1997) suggest that it should. The ability of transposed stimuli to convey both simple (pure-tone) and complex (multitone) pitch was investigated by Oxenham et al. (2004). Using frequency-discrimination and pitch-matching tasks, they found that simple pitch perception was poorer with transposed stimuli than with pure-tone stimuli. Perhaps more importantly, they found that transposing the third through the fifth harmonics to different high-frequency carriers failed to elicit a pitch percept at all.

This finding suggests that, in apparent contrast to low-frequency binaural processing, pitch processing is sensitive to the tonotopic location of the temporal information in the cochlea. This does not necessarily imply a place code for pitch, but it does suggest that the way temporal information is evaluated depends on its spatial location in the cochlea. There are a number of possible explanations of this finding. One is that only a certain range of interspike intervals is evaluated at each characteristic frequency (Moore 1982). Another is that fine structure and envelope are encoded in fundamentally different ways. This could occur if the fine structure were encoded via the rapid phase transitions around CF (Shamma 1985a; Shamma and Klein 2000), whereas the envelope were encoded via a mechanism based more directly on timing intervals between successive neural events.

### *4.3 Separate Pitch Mechanisms for Resolved and Unresolved Harmonics?*

Many modern models of pitch suggest that the auditory system performs a synthesis of the information from the resolved and unresolved harmonics in order to derive a final pitch. Furthermore, they propose that the same algorithm is used to extract information about F0 from the two harmonic groups. Some recent psychophysical evidence, however, suggests that the auditory system may process the information from resolved and unresolved harmonics in different ways. It has been argued that there may be two pitch mechanisms, one for resolved harmonics, and one for unresolved harmonics.

There is a certain amount of indirect evidence for this claim. First, F0 discrimination is much worse for unresolved than for resolved harmonics (Houtsma and Smurzynski 1990), even when the discriminations are made in the same spectral region (see Fig. 2.7; Shackleton and Carlyon 1994). Second, the improvement in F0 discrimination with tone duration is greater for unresolved than for resolved harmonics, suggesting that different integration mechanisms may be involved (Plack and Carlyon 1995; see Section 6.1.3).

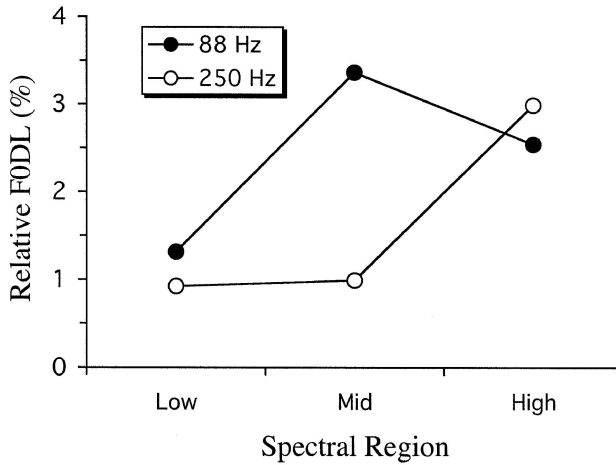


FIGURE 2.7. The results of Shackleton and Carlyon (1994) showing the F0DL (as a percentage of F0) as function of F0 (shown in the legend) and spectral region. For each F0, harmonics were filtered into one of three spectral regions, low (125–625 Hz), mid (1375–1875 Hz), and high (3900–5400 Hz). The harmonics of the 88-Hz complex were resolved in the low region but unresolved in the mid and high regions. The harmonics of the 250-Hz complex were resolved in the low and mid regions, but unresolved in the high region. The results for the mid region show that discrimination performance is worse for a group of unresolved harmonics, even when they occupy the same spectral region as a group of resolved harmonics.

A study by Carlyon and Shackleton (1994) suggested that the pitches from resolved and unresolved harmonics may involve different *encoding* mechanisms. Carlyon and Shackleton (1994) presented *simultaneously* two groups of harmonics with the same nominal F0 (either 88 or 250 Hz) that were filtered into two separate spectral regions, chosen from “low” (125 to 625 Hz), “mid” (1375 to 1875 Hz), and “high” (3900 to 5400 Hz). A “dynamic” F0 difference between the groups was introduced by frequency modulating their F0s 180° out of phase. When the combination of F0 and spectral regions was such that one group of harmonics was resolved and the other was unresolved (e.g., 88-Hz low, which contains resolved harmonics, versus 88-Hz mid, which contains unresolved harmonics), then F0 discrimination between the groups was poor compared to situations in which both groups were resolved (250-Hz low versus 250-Hz mid) or in which both groups were unresolved (88-Hz mid versus 88-Hz high). The unresolved versus unresolved comparison was probably mediated by the detection (across frequency) of asynchronies between the envelope peaks of the two groups during the course of the modulation (“pitch pulse asynchronies”). This is a cue that does not depend on an extraction of F0. However, using an analysis based on signal detection theory, Carlyon and Shackleton showed that the *simultaneous* resolved versus unresolved F0 discriminations were worse than would be expected on the basis of resolved versus resolved and unresolved

versus unresolved F0 discriminations in a *sequential* task (i.e., no pitch-pulse asynchronies). They argued that there is an additional difficulty (“translation noise”) when making comparisons between resolvability groups, as the pitches are encoded by separate mechanisms. On the other hand, Gockel et al. (2004) showed that when a resolved and an unresolved group of harmonics, with similar F0s but filtered into different spectral regions, are presented simultaneously, the resolved group dominates the pitch percept and *interferes* with pitch processing for the unresolved group. This could provide an alternative explanation for why simultaneous resolved versus unresolved comparisons are difficult. Recent evidence suggests that *sequential* F0 comparisons between resolved and unresolved groups do not exhibit translation noise (Micheyl and Oxenham 2004).

Finally, Grimault et al. (2002) examined the effects of selective training on F0 discrimination. They first tested all their listeners on F0 discrimination for groups of resolved and unresolved harmonics (using the same combinations of F0 and spectral region employed by Carlyon and Shackleton 1994: 88-Hz low, mid, and high; and 250-Hz low, mid, and high). They then divided the listeners into three groups. One group was trained over a period of 4 weeks on F0 discrimination with a specific group of resolved harmonics (250-Hz mid), one group was trained with a specific group of unresolved harmonics (88-Hz mid), and a control group received no training. After 4 weeks they retested the listeners with the first set of conditions. Although both the trained groups performed better than the control group on all the conditions, listeners trained with resolved harmonics showed a greater improvement in performance on *all* the resolved conditions (88-Hz low, 250-Hz low and mid) than they did on the unresolved conditions (88-Hz mid and high, 250-Hz high), and vice versa. In other words, the effect of training was specific to the resolvability of the components to some extent. Again, this suggests that separate mechanisms are involved in encoding the pitches of resolved and unresolved harmonics.

The evidence for two pitch mechanisms tallies with the idea that the pitch of a resolved complex may be derived from a template matched to the individual harmonics (Goldstein 1973; Terhardt 1974), whereas the pitch of an unresolved complex may be derived by a purely temporal mechanism, operating on the interaction of the higher harmonics on the basilar membrane (Schouten 1970). However, the fact that the pitches of resolved and unresolved harmonics can be compared, and can be used to produce similar percepts (e.g., musical melodies), suggests that there is a convergence of the representations at some stage in the auditory pathway.

## 5. Dichotic Pitch

The term dichotic pitch refers to situations in which two noises, which individually produce no pitch, elicit a pitch sensation when presented simultaneously to opposite ears. The effect has been likened to random-dot stereograms in vision (Julesz 1971), in that the percept requires semicoherent (or partially cor-

related) input to both ears (or eyes) to emerge (Akeroyd et al. 2001). The first such pitch to be described has come to be known as Huggins pitch (Cramer and Huggins 1958). This pitch is produced by introducing a rapid but smooth phase transition within a narrow spectral region of an otherwise binaurally coherent noise (see Fig. 2.8, left panel). Another pitch that has received considerable attention is the binaural edge pitch (Klein and Hartmann 1981), which involves two noises, one in each ear, which are in phase below a certain frequency and out of phase above that frequency (Fig. 2.8, middle panel). A more recent, but related, addition to the family of dichotic pitches is the binaural coherence edge pitch (Hartmann and McMillon 2001), where the cutoff frequency marks the transition between correlated and uncorrelated noise (Fig. 2.8, right panel). A second class of dichotic pitches has been termed “Fourcin pitch” and involves the simultaneous binaural presentation of different independent noises to the two ears, with each noise associated with a different interaural time delay (Fourcin 1970; Bilsen and Goldstein 1974). If there are two noises and one of them has an interaural phase shift of 180 degrees, the perceived periodicity corresponds to the difference in the interaural delays between the two noises.

Akeroyd et al. (2001) tested listeners’ abilities to use dichotic pitch to recognize well-known melodies with all rhythmic information removed. Using Huggins pitch, binaural-edge pitch and binaural coherence edge pitch, they found that all three stimuli produced a sufficiently strong pitch to carry melodic information and that performance was good even in the first block of trials, showing that extended exposure or practice is not necessary to hear dichotic pitches. However, there was a clear hierarchy in their results: overall the Huggins pitch produced the most salient pitch (as evidenced by better melody rec-

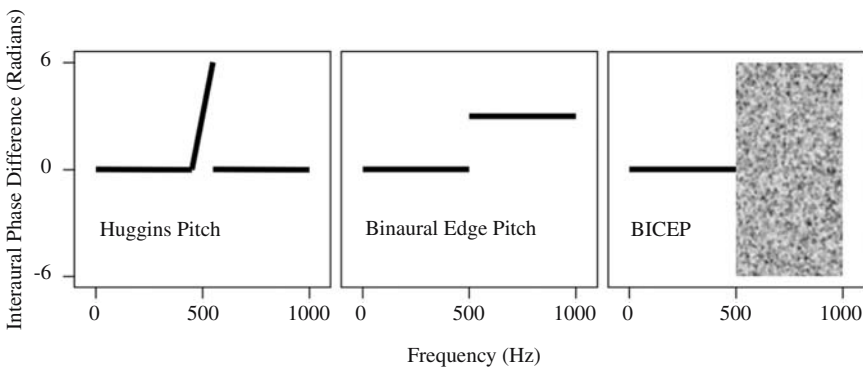


FIGURE 2.8. A schematic illustration of three different binaural pitch stimuli: Huggins pitch, binaural edge pitch, and binaural coherence edge pitch (BICEP). The figure plots the phase difference between a wideband noise presented to the left ear and a wideband noise presented to the right ear, as a function of frequency. The figure is based on Figure 1 in Akeroyd et al. (2001).

ognition), with the binaural-edge pitch producing similar, but slightly poorer results. The binaural coherence edge pitch produced somewhat poorer results, although still well above chance.

Dichotic pitches have been used to test models of binaural perception (Culling et al. 1998a,b; Culling 2000), but also have some relevance for models of pitch in general. In particular, the findings provide evidence that pitch can be formed centrally and that neither monaural spectral nor monaural temporal information is necessary to elicit a pitch sensation that can be used by listeners to follow a melody.

## 6. Temporal Integration

Any measure of repetition rate or frequency has to be obtained over a certain duration, since these quantities are defined in terms of patterns of activity *over time*. The questions are: What integration mechanism does the auditory system use to derive pitch and how is information combined over time to improve the accuracy of the pitch estimate? In the integration of intensity or loudness, it seems likely that very different integration times are used by the auditory system for tasks that require the detection of rapid changes in intensity (e.g., gap detection) and for tasks that may be aided by a long accumulation of information over time (e.g., detection of long-duration tones in noise). Similarly it may be necessary to distinguish between the minimum integration time of the pitch mechanism, which determines our ability to follow rapid changes in frequency or F0, and a long integration time that may be used in frequency or F0 discrimination tasks with long-duration tones.

### 6.1 Measures of Temporal Integration

#### 6.1.1 Sensitivity to Modulation

The ability of listeners to follow the pitch of a stimulus with a varying repetition rate (vibrato) provides important information about the minimum integration time of the pitch mechanism. If the mechanism is sluggish (long integration time) then the individual fluctuations will be averaged together and the system will not be sensitive to the modulation.

Unfortunately, modulating the frequency of a pure tone or the F0 of a group of resolved harmonics will tend to induce *amplitude* modulations because of the characteristics of cochlear filtering: sweeping a component across a highly tuned bandpass filter will result in fluctuations in amplitude at the output of the filter. It is known that the auditory system is very sensitive to amplitude modulation (AM; see Viemeister 1979). For moderate modulation rates, above 5 to 10 Hz, or for carrier frequencies above about 4000 Hz, it is likely that the detection of sinusoidal frequency modulation (FM) for pure-tone carriers is based on the excitation-pattern cues associated with induced AM (Zwicker and Fastl 1990;



Moore and Sek 1994, 1996). For even higher modulation rates, the FM will be detected by the presence of resolved spectral sidebands. For very low modulation rates, detection may be based on following the changes in phase locking as the frequency changes (i.e., a temporal pitch mechanism). Sek and Moore (1995) argued that the decrease in sensitivity to FM with increasing modulation rate (over the range from 2 to 10 Hz) suggests that the mechanism that decodes the phase-locking information is sluggish. They pointed out that for a 2-Hz modulation rate, the instantaneous frequency of the pure tone is within 10% of the frequency extremes for around 70 ms each cycle. The corresponding figure for 5-Hz FM is around 30 ms. The DLF increases dramatically over this range of durations (see Fig. 2.1).

Modulating the F0 of a group of unresolved harmonics, while passing the components through a fixed bandpass filter, avoids the problems of induced AM and sideband detection. The amplitude at the output of the auditory filters will change very little as a result of variation in the frequencies of the individual harmonics, because several harmonics fall within each filter. Although there will be a slight induced AM produced by variations in the spacing of harmonics as F0 is varied, for small FM depths this should not be detectable. Plack and Carlyon (1995) showed that listeners were much worse at detecting 5-Hz sinusoidal F0 modulation of complex tones with unresolved harmonics (threshold depth around 10%), than of complex tones with resolved harmonics (threshold depth around 0.5%). They argued that this was because the pitch mechanism for unresolved harmonics needs a long duration in order to make an accurate estimate of F0. In a more comprehensive study, Carlyon et al. (2000) measured the detection of F0 modulation as a function of modulation rate. For both resolved and unresolved harmonics, the modulation depth at threshold increased with modulation rate for rates above 2 Hz. Again, this low-pass characteristic suggests that the pitch mechanism requires a long duration to make an accurate estimate of F0, and that rapid fluctuations in F0 may be essentially “averaged out” by the integration window.

When the FM rate and depth are not too high, a single pitch may be assigned to a modulated complex tone (d’Alessandro and Castellengo 1994) or pure tone (Gockel et al. 2001). Gockel et al. (2001) obtained pitch matches between an unmodulated pure tone with an adjustable frequency and a pure tone (frequency 500 to 8000 Hz) that was frequency modulated (rate 5 to 20 Hz, depth 8%) according to a repeated U pattern (UU, etc.) or inverted U pattern. In other words, the instantaneous frequency changed very rapidly, except in the middle of each repetition (the bowl of the U), where the change was slower. They found that the matched frequency was shifted away from the mean frequency of the modulation toward the portion of the modulation that had the slowest rate of change (i.e., a downward shift for the U pattern and an upward shift for the inverted U pattern). Gockel et al. (2001) argued that the overall pitch of a frequency-modulated sound corresponds to a weighted average of individual estimates of the period, with lower weights given to the estimates obtained during rapid changes in period. They also argued that the weight given should

be related directly to a compressive function of the amplitude of the waveform at each time. Earlier models, based on the envelope-weighted average of instantaneous frequency (EWAIF) or the intensity-weighted average of instantaneous frequency (IWAIF) (Feth 1974; Feth et al. 1982), did not provide a good description of the pitch shifts observed by Gockel et al. (2001).

### 6.1.2 Pitch Fluctuations in Repeated Period Noise

Repeated period noise is generated by concatenating noise samples of equal duration. When the sequence contains consecutive noise samples that are identical (e.g., AABBCDD, etc.), a periodicity is generated in the waveform that can be heard as a pitch corresponding to  $1/d$  kHz, where  $d$  is the duration of the noise sample in milliseconds. Wiegand (2001) modified this stimulus by inserting independent noise samples between sequences of identical noise samples (e.g., AAABCDEEEFGH, etc.). In this way, he was able to vary the duration of uncorrelated noise independently of  $d$ . If the rate of oscillation between periods of identical noise samples and periods of independent noise samples was slow enough, then the oscillations could be heard as regular fluctuations in pitch strength. If the rate of oscillation was too high, however, then little variation in pitch strength was heard, presumably because the periods of correlation and independence were being averaged together by the auditory system. By measuring the salience of the pitch-strength fluctuations as a function of the rate of oscillation, Wiegand was able to estimate time constants for the pitch integration process, based on an analysis using an ACF model. In this case the time constant is that of an exponential integrator, integrating correlation strength over time. He found that the time constant increased (longer integration) with decreasing repetition rate ( $1/d$ ), suggesting that the time constant may be dependent on the autocorrelation delay in the ACF. In other words, different time constants may be used at different delays for the production of a single ACF. Wiegand estimated that the time constants were about 2.5 ms for delays less than 1.25 ms, and double the delay for delays greater than 1.25 ms. So for a 250-Hz periodicity ( $d = 4$  ms), the time constant would be about 8 ms for the first peak in the ACF, and about 16 ms for the second peak in the ACF. It must be emphasized, however, that these are probably estimates of the *minimum* integration time of the system.

### 6.1.3 Effects of Duration on Discrimination

Frequency and F0 discrimination improve with duration, and the time course of this improvement may provide an estimate of the long (or possibly maximum) integration time(s) for pitch. The effect is much greater for complex tones with unresolved harmonics and for pure tones than for complex tones with resolved harmonics with the same repetition rate. For an F0 of 250 Hz, White and Plack (1998) found little improvement in performance beyond 40 ms for resolved harmonics, but improved performance out to a duration of 80 ms for unresolved harmonics (Fig. 2.9). Just as the effect of duration on the FDL for pure tones

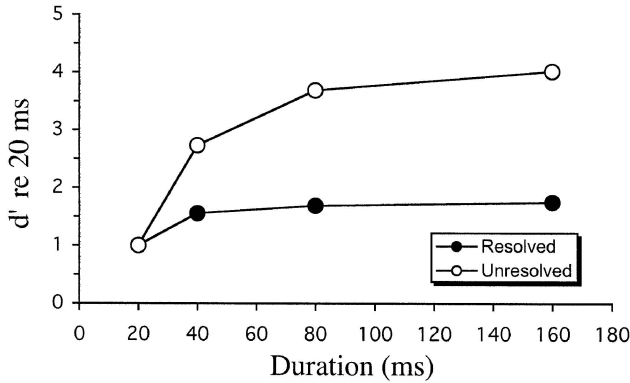


FIGURE 2.9. The F0 discrimination results of White and Plack (1990), showing the detectability index,  $d'$ , as a function of duration for groups of resolved and unresolved harmonics. The value for  $d'$  is plotted relative to the value for the 20-ms complex for each group. For each harmonic group, the F0 difference between the two complexes being compared was fixed across the different durations, and  $d'$  was derived from the percent correct discrimination.

increases with decreasing frequency (see Section 2.2), so the effect of duration on the F0DL for unresolved harmonics increases with decreasing F0: for a 62.5-Hz complex, White and Plack found clear improvements with duration up to a duration of 160 ms (the longest duration they used). Consistent with the interpretation of Wiegand (2001), this may mean that the integration time is longer for low F0s.

Plack and Carlyon (1995) noted that the improvement in performance with increasing duration for unresolved harmonics was similar to that for a pure tone with a frequency equal to the F0 of the complex. The improvement for *resolved* harmonics, however, was similar to that for a pure tone with a frequency close to the dominant region of the complex. They suggested that the auditory system may determine the individual frequencies of the resolved harmonics, but process only the overall repetition rate of the unresolved harmonics, not making full use of the fine structure information.

This observation may have some relevance for models of pitch perception. A pitch mechanism that simply examines the interspike intervals equal to  $1/F_0$  across channels (such as the summary ACF model of Meddis and Hewitt 1991; and the schematic model described by Moore 2003) may not be making optimal use of the temporal information present in the auditory nerve. For example, such a mechanism would ignore the interspike intervals of 5 ms produced by the 2nd harmonic of a 100-Hz F0, and process only the 10-ms interspike intervals. However, the 5-ms intervals are providing information that constrains the range of possible F0s, and this information should not be discarded by an optimal processor. The fact that discrimination performance is very good for com-

plexes with resolved harmonics consisting of only five waveform cycles, in contrast to pure tones and complex tones with unresolved harmonics, suggests that more information is being extracted from the resolved harmonics than is suggested by some pitch models.

6.1.4 Integration of Nonsimultaneous Harmonics

Ciocca and Darwin (1999) found that a mistuned 4th harmonic in a complex tone contributes to the pitch of the complex as a whole even when it is presented before or after the rest of the complex (see Fig. 2.10). Indeed, if the harmonic is presented after the complex, the mistuning is still effective when the silent gap between the complex and the harmonic is 80 ms. The work is an extension of an earlier study by Hall and Peters (1981), who showed that three successive harmonics (with a total duration of 140 ms) could be integrated together to form a unified pitch, if they were presented in a background noise at a low signal-to-noise ratio (see Section 3.5.3). Similarly, Grose et al. (2002) measured F0 discrimination for groups of sequentially presented 40-ms harmonics (i.e., no two harmonics present simultaneously) in background noise. They found that, for harmonic separations up to about 45 ms, performance was almost as good as that for synchronous harmonics. Since in their design at least three harmonics must have been included to produce a reliable estimate of F0, Grose et al. suggested a minimum integration time of around 210 ms.

These results suggest that the integration time for resolved harmonics is much longer than may be expected on the basis of the F0 discrimination data described in Section 6.1.3. However, the fact that discrimination performance for continuous tones does not improve as duration is increased, does *not* imply that the

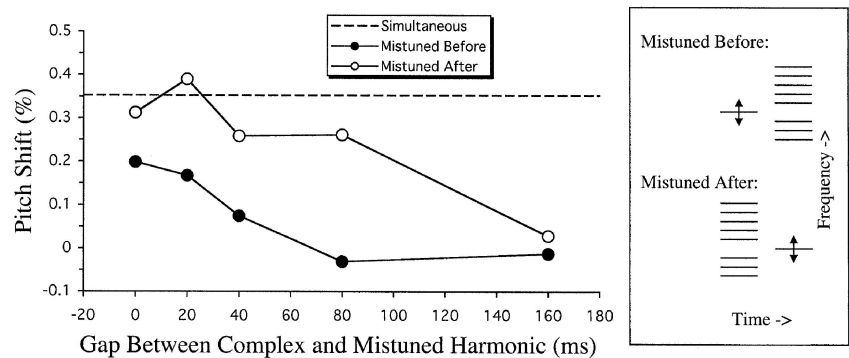


FIGURE 2.10. The results of Ciocca and Darwin (1999) showing the shift in periodicity pitch produced by mistuning the fourth harmonic of a complex tone by 3%, as a function of the silent interval between the mistuned harmonic and the rest of the complex tone. The mistuned harmonic was presented either before or after the rest of the complex (see schematic spectrogram on the right).

pitch estimate is based on only a short integration time. It is possible that long integration occurs, but does not contribute to the *accuracy* of the pitch estimate in some cases. For example, there could be a central limitation that puts a cap on performance. Once performance has improved to this level, further increases in duration may have no effect on performance. Another possibility is that the auditory system may vary the integration time depending on the demands of the task. For example, the integration time may be increased if temporally disparate information needs to be combined to produce a pitch estimate.

## 6.2 Integration Mechanisms

### 6.2.1 Multiple Looks and Long Integration Windows

Performance on psychophysical tasks is often assumed to be limited by “internal noise,” that is, variability in the internal (neural) representation of stimuli. If such a noise is independent from one time to the next, then performance can be improved by simply adding, or averaging, the overall activity across time. In the formulation of Green and Swets (1966), the detectability ( $d'$ ) of a change in a stimulus is dependent on the magnitude of the internal representation of the change divided by the standard deviation of the internal noise. If several samples, or “looks” (Viemeister and Wakefield 1991), are added, then the internal representation of the change will increase linearly with the number of looks, whereas the standard deviation of the representation will increase according to the *square root* of the number of looks. This is just a property of adding random Gaussian variables. It follows that if the number of samples of a stimulus increases by a factor  $n$ , then  $d'$  should increase by a factor  $\sqrt{n}$ . This assumes, of course, that the auditory system can make optimal use of the information.

It is important to realize that the information need not be combined continuously for a multiple-looks strategy to work. Viemeister and Wakefield (1991) showed that the detection threshold for two brief tone bursts separated in time was lower than that for a single burst, and was not affected by changes in the level of an intervening masking noise. They suggested that the auditory system sampled the two tones *discretely* using short integration windows, and combined the information optimally at a later stage.

White and Plack (1998) found evidence for a similar process operating in the pitch domain. In one condition, they presented two pairs of 20-ms tone bursts, containing either resolved or unresolved harmonics of a 250-Hz F0, with the members of each pair separated by a brief gap of 5 ms or more. The F0 of one pair was higher than that of the other, and listeners were required to indicate which pair had the higher pitch. In another condition, they required listeners to compare the pitches of two single 20-ms tone bursts. White and Plack used the method of constant stimuli for this experiment (see Section 1.1) with a fixed F0 difference for the resolved harmonics and a fixed F0 difference for the unresolved harmonics. For both resolved and unresolved harmonics, they found that the  $d'$  for a pair of bursts was a factor of  $\sqrt{2}$  greater than the  $d'$  for a single

burst, as predicted by the multiple-looks hypothesis. Significantly, performance did not change as the time interval between the tone bursts in a pair was increased. One interpretation of these data is that the auditory system is able to estimate the F0s of the two tone bursts discretely, and combine these samples to produce a more reliable final estimate. The results of Gockel et al. (2001) for modulated pure tones (Section 6.1.1) suggest that, when combining the individual samples, the auditory system may weight samples taken when the period of the waveform is changing slowly more highly than samples taken when the period is changing rapidly.

The large improvement with duration in F0 discrimination for unresolved harmonics (and in the FDL for low-frequency pure tones), however, suggests that the pitch mechanism for these stimuli combines information over time in a way that goes beyond the combination of independent discrete samples. For example,  $d'$  for the detection of an F0 difference for a 62.5-Hz complex with unresolved harmonics increases by a factor of three as duration is increased from 40 to 80 ms (White and Plack 1998). This is much greater than the  $\sqrt{2}$  improvement predicted by the multiple-looks model, and suggests that some other process may be involved. This process may include a long integration window, which means that over a certain continuous duration, the neural activity produced by the stimulus is analyzed together in some way. A possible analogue is the discrete Fourier transform (DFT). If a continuous pure tone is sampled using a short window before the DFT is calculated, then the spectral representation is not as sharp as it would be if a long sampling window were used. This is just a consequence of the *time–frequency tradeoff* (see de Cheveigné, Chapter 6). Although it is unlikely that the auditory system performs a DFT on neural activity, the principle is the same: the longer the time window the more accurate (potentially) the estimate of periodicity. In terms of common-interval mechanisms such as the ACF, using long-term periodicity would require an analysis of high-order intervals between spikes (e.g., Heinz et al. 2001a).

As described earlier, the duration of the integration window may depend on the demands of the task. When the task involves following rapid changes in frequency or F0 (such as in an FM detection task) listeners may use a short integration window to maximize temporal acuity. For frequency and F0 discrimination between static tones, a long window might be used to maximize the information falling within the window. It is hard to specify the duration and shape of the window because we do not have a clear idea about the quantity that is being integrated (possibilities include correlation strength or number of pulses) or how changes in the amount of integration relate to changes in performance. It does seem likely, however, that integration windows with durations in excess of 100 ms may be used by the auditory system in some circumstances.

### 6.2.2 Resetting and Continuity

In their experiments on the integration mechanism for unresolved harmonics (see Section 6.2.1), White and Plack (1998) observed that inserting a gap of only 5 ms between two 20-ms bursts of a 250-Hz complex was sufficient to

produce a large deterioration in F0 discrimination. They argued that if the auditory system were using a fixed long integration time, then the presence of the gap should have little effect on performance. The fact that the improvement in performance from one burst to two was consistent with a multiple looks mechanism when there was a gap between the bursts, suggests that the auditory system may use a *flexible* integration time for pitch. For continuous tones a long integration time may be used, but the integration time is reset, to start a new F0 estimate, in response to temporal discontinuities. Such a resetting mechanism may be useful in the environment, where a temporal discontinuity often reflects the end of one auditory object and the beginning of another one. It would be appropriate to analyze the F0s of these objects separately.

A similar conclusion was reached by Nabelek (1996) regarding pure tones. He presented two tone bursts separated by a gap. The tone bursts started and ended with zero phase. When the gap was not an integer number of periods (e.g., if the frequency was 1250 Hz and the gap was 2 ms) a phase difference existed between the phase of the second burst and the phase the first burst would have had if it were continuous. In these situations, Nabelek observed a shift in the pitch of the tone burst pair, relative to the nominal frequency. However, this occurred only when the gap between the two bursts was less than a “critical pause duration” of between 8 and 16 ms. For gaps larger than this, the two tone bursts appeared to be processed separately, and the relative phases of the bursts had no effect. Previously Bregman et al. (1994a,b) had suggested that the *onset* of a pure tone may cause the auditory system to reset and begin a new frequency estimate.

Plack and White (2000b) wondered whether the hypothetical resetting mechanism is sensitive to “illusory” continuity. If the gaps in a tone are filled by a noise, sufficient to mask the tone when presented simultaneously, then the tone is perceived as being continuous (e.g., Elfner and Caskey 1965; Houtgast 1973). The auditory system, quite sensibly, interprets the noise as an extraneous sound superimposed on a continuous tone. If the resetting mechanism is used to allow separate analysis of different auditory objects, then the mechanism may not operate when the perceptual evidence suggests that the two tone bursts belong to the same auditory object. In line with this prediction, when Plack and White inserted a noise in the gap between the two bursts to produce a perception of continuity, F0 discrimination improved to the level observed when the tone bursts really were continuous. It should be noted that there may be more prosaic explanations for some of these findings. For example, the auditory nervous system is very sensitive to stimulus onsets, and it is possible that the extra onset produced by adding a silent gap may disrupt performance in some way.

## 7. Summary

The psychophysical results described in this chapter suggest that pitch is a very complicated percept. A wide range of stimuli from pure tones, through harmonic complex tones, amplitude-modulated and iterated noises, to stimuli based

on interaural correlation, all produce a sensation of pitch. It seems that the auditory system is extremely sensitive to stimulus regularity. Furthermore, manipulations of these stimuli by varying the frequencies of harmonics, the timing of pitch pulses, and the temporal and spectral envelopes, produce a range of effects that provide severe tests for physiological theories and computational models of pitch perception. As some authors have argued, it may be impossible to account for all these effects by a mechanism that does not explicitly differentiate between lower and higher harmonics.

Some psychophysical experiments have been criticized by those who can see little connection between the esoteric stimuli used in the laboratory and the tones that we hear in our environment (or perhaps the tones that would have been important during the evolution of the human, or mammalian, auditory system). However, it is often necessary to resort to esoteric-seeming stimuli in order to discern the mechanisms by which the pitches of more usual stimuli are coded. Also, most of the stimuli described in this chapter can be shown to have a *musical* pitch, in that they can be used to produce melodies. Either the auditory system learns to label these sensations as musical during the course of the experiment, or there is at least some connection between the sensations produced by unresolved complex tones, modulated noise, IRN, and so forth and the pitch that is produced by more “realistic” tones such as wideband complexes.

So how should we summarize such a diverse set of findings? First, the simplest theories about pitch that may apply to a strictly periodic stimulus often fail when the periodicity and harmonicity of a waveform is disrupted. Considering a range of experimental results, it is clear that pitch is not a simple function of waveform repetition rate or of harmonic spacing. Second, the processing limitations of the peripheral auditory system have important consequences for pitch perception, in terms of the spectral resolution of individual harmonics, and also perhaps in terms of the relation of the existence regions of pitch to the limits of phase locking in the auditory nerve. Finally, and on a positive note, we can confidently state that progress is being made. A casual glance at the bibliography reveals how much important research has been done over the last few years. We know much more about the psychophysics of pitch than we did a decade ago, and with the current increased interest in the field there is every reason to be optimistic for the future.

*Acknowledgments.* Many thanks to Brian Moore, Dick Fay, Christophe Micheyl, Josh Bernstein, and John Culling for detailed comments on an earlier version of this chapter. Thanks also to Roy Patterson for advice on iterated rippled noise, to Daniel Pressnitzer for discussions on autocorrelation modeling, and to Michael Akeroyd for help with Figure 2.8. The authors receive support for their work from the Engineering and Physical Sciences Research Council (GR/R65794/01 to C.J. Plack) and the National Institutes of Health (R01 DC 05216 to A.J. Oxenham).



## References

- Akeroyd MA, Moore BCJ, Moore GA (2001) Melody recognition using three types of dichotic-pitch stimulus. *J Acoust Soc Am* 110:1498–1504.
- Arehart KH, Burns EM (1999) A comparison of monotic and dichotic complex-tone pitch perception in listeners with hearing loss. *J Acoust Soc Am* 106:993–997.
- Attneave F, Olson RK (1971) Pitch as a medium: a new approach to psychophysical scaling. *Am J Psychol* 84:147–166.
- Berg BG (1989) Analysis of weights in multiple observation tasks. *J Acoust Soc Am* 86:1743–1746.
- Bernstein JG, Oxenham AJ (2003) Pitch discrimination of diotic and dichotic complexes: harmonic resolvability or harmonic number? *J Acoust Soc Am* 113:3323–3324.
- Bernstein LR, Trahiotis C (2002) Enhancing sensitivity to interaural delays at high frequencies by using “transposed stimuli.” *J Acoust Soc Am* 112:1026–1036.
- Bilsen FA, Goldstein JL (1974) Pitch of dichotically delayed noise and its possible spectral basis. *J Acoust Soc Am* 55:292–296.
- Bregman AS, Ahad PA, Kim J (1994a) Resetting the pitch-analysis system. 2. Role of sudden onsets and offsets in the perception of individual components in a cluster of overlapping tones. *J Acoust Soc Am* 96:2694–2703.
- Bregman AS, Ahad P, Kim J, Melnerich L (1994b) Resetting the pitch-analysis system: 1. Effects of rise times of tones in noise backgrounds or of harmonics in a complex tone. *Percept Psychophys* 56:155–162.
- Burns EM, Viemeister NF (1976) Nonspectral pitch. *J Acoust Soc Am* 60:863–869.
- Burns EM, Viemeister NF (1981) Played again SAM: further observations on the pitch of amplitude-modulated noise. *J Acoust Soc Am* 70:1655–1660.
- Cariani PA, Delgutte B (1996) Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *J Neurophysiol* 76:1698–1716.
- Carlyon RP (1996) Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker. *J Acoust Soc Am* 99:517–524.
- Carlyon RP (1997) The effects of two temporal cues on pitch judgements. *J Acoust Soc Am* 102:1097–1105.
- Carlyon RP, Shackleton TM (1994) Comparing the fundamental frequencies of resolved and unresolved harmonics: evidence for two pitch mechanisms? *J Acoust Soc Am* 95:3541–3554.
- Carlyon RP, Moore BC, Micheyl C (2000) The effect of modulation rate on the detection of frequency modulation and mistuning of complex tones. *J Acoust Soc Am* 108:304–315.
- Carlyon RP, van Wieringen A, Long CJ, Deeks JM (2002) Temporal pitch mechanisms in acoustic and electric hearing. *J Acoust Soc Am* 112:621–633.
- Ciocca V, Darwin CJ (1999) The integration of nonsimultaneous frequency components into a single virtual pitch. *J Acoust Soc Am* 105:2421–2430.
- Cramer EM, Huggins WH (1958) Creation of pitch through binaural interaction. *J Acoust Soc Am* 30:413–417.
- Culling JF (2000) Dichotic pitches as illusions of binaural unmasking. III. The existence region of the Fourcin pitch. *J Acoust Soc Am* 103:3509–3526.
- Culling JF, Summerfield AQ, Marshall DH (1998a) Dichotic pitches as illusions of binaural unmasking. I. Huggins’ pitch and the “binaural edge pitch.” *J Acoust Soc Am* 103:3509–3526.

- Culling JF, Marshall DH, Summerfield AQ (1998b) Dichotic pitches as illusions of bin-aural unmasking. II. The Fourcin pitch and the dichotic repetition pitch. *J Acoust Soc Am* 103:3527–3539.
- Dai H (2000) On the relative influence of individual harmonics on pitch judgment. *J Acoust Soc Am* 107:953–959.
- d'Alessandro C, Castellengo M (1994) The pitch of short-duration vibrato tones. *J Acoust Soc Am* 95:1617–1630.
- Darwin CJ (1992) Listening to two things at once. In: Schouten MEH (ed), *The Auditory Processing of Speech: From Sounds to Words*. Berlin: Mouton de Gruyter, pp. 133–147.
- de Cheveigné A (1999) Pitch shifts of mistuned partials: a time-domain model. *J Acoust Soc Am* 106:887–897.
- Elfner LF, Caskey WE (1965) Continuity effects with alternating sounded noise and tone signals as a function of manner of presentation. *J Acoust Soc Am* 38:543–547.
- Emmerich DS, Ellermeier W, Butensky B (1989) A re-examination of the frequency discrimination of random-amplitude tones, and a test of Henning's modified energy-detector model. *J Acoust Soc Am* 85:1653–1659.
- Faulkner A (1985) Pitch discrimination of harmonic complex signals: residue pitch or multiple component discriminations. *J Acoust Soc Am* 78:1993–2004.
- Feth LL (1974) Frequency discrimination of complex periodic tones. *Percept Psychophys* 15:375–379.
- Feth LL, O'Malley H, Ramsey JJ (1982) Pitch of unresolved, two-component complex tones. *J Acoust Soc Am* 72:1403–1412.
- Flanagan JL, Guttman N (1960) On the pitch of periodic pulses. *J Acoust Soc Am* 32:1308–1319.
- Fourcin AJ (1970) Central pitch and auditory lateralization. In: Plomp R, Smoorenburg GF (eds), *Frequency Analysis and Periodicity Detection in Hearing*. Leiden: Sijthoff, pp. 319–328.
- Glasberg BR, Moore BCJ (1990) Derivation of auditory filter shapes from notched-noise data. *Hear Res* 47:103–138.
- Gockel H, Moore BCJ, Carlyon RP (2001) Influence of rate of change of frequency on the overall pitch of frequency-modulated tones. *J Acoust Soc Am* 109:701–712.
- Gockel H, Carlyon RP, Plack CJ (2004) Across frequency interference effects in fundamental frequency discrimination: questioning evidence for two pitch mechanisms. *J Acoust Soc Am* 116:1092–1104.
- Goldstein JL (1973) An optimum processor theory for the central formation of the pitch of complex tones. *J Acoust Soc Am* 54:1496–1516.
- Green DM, Swets JA (1966) *Signal Detection Theory and Psychophysics*. New York: Krieger.
- Grimault N, Micheyl C, Carlyon RP, Collet L (2002) Evidence for two pitch encoding mechanisms using a selective auditory training paradigm. *Percept Psychophys* 64:189–197.
- Grose JH, Hall JW, Buss E (2002) Virtual pitch integration for asynchronous harmonics. *J Acoust Soc Am* 112:2956–2961.
- Haftner ER, Saberi K (2001) A level of stimulus representation model for auditory detection and attention. *J Acoust Soc Am* 110:1489–1497.
- Hall JW, Peters RW (1981) Pitch from nonsimultaneous successive harmonics in quiet and noise. *J Acoust Soc Am* 69:509–513.
- Hall JWI, Buss E, Grose JH (2003) Modulation rate discrimination for unresolved com-

- ponents: temporal cues related to fine structure and envelope. *J Acoust Soc Am* 113: 986–993.
- Hartmann WM (1997) *Signals, Sound, and Sensation*. New York: Springer-Verlag.
- Hartmann WM, Doty SL (1996) On the pitches of the components of a complex tone. *J Acoust Soc Am* 99:567–578.
- Hartmann WM, McMillon CD (2001) Binaural coherence edge pitch. *J Acoust Soc Am* 109:294–305.
- Hartmann WM, McAdams S, Smith BK (1990) Hearing a mistuned harmonic in an otherwise periodic complex tone. *J Acoust Soc Am* 88:1712–1724.
- Heinz MG, Colburn HS, Carney LH (2001a) Evaluating auditory performance limits: I. One-parameter discrimination using a computational model for the auditory nerve. *Neural Comput* 13:2273–2316.
- Heinz MG, Colburn HS, Carney LH (2001b) Evaluating auditory performance limits: II. One-parameter discrimination with random-level variation. *Neural Comput* 13:2317–2338.
- Helmholtz HLF (1863) *Die Lehre von den Tonempfindungen als Physiologische Grundlage für die Theorie der Musik*. Braunschweig: F. Vieweg.
- Henning GB (1966) Frequency discrimination of random amplitude tones. *J Acoust Soc Am* 39:336–339.
- Houtgast T (1973) Psychophysical experiments on “tuning curves” and “two-tone inhibition.” *Acustica* 29:168–179.
- Houtgast T (1976) Subharmonic pitches of a pure tone at low S/N ratio. *J Acoust Soc Am* 60:405–409.
- Houtsma AJM (1995) Pitch perception. In: Moore BCJ (ed), *Hearing*. Orlando, FL: Academic Press, pp. 267–295.
- Houtsma AJM, Goldstein JL (1972) The central origin of the pitch of pure tones: evidence from musical interval recognition. *J Acoust Soc Am* 51:520–529.
- Houtsma AJM, Smurzynski J (1990) Pitch identification and discrimination for complex tones with many harmonics. *J Acoust Soc Am* 87:304–310.
- Johnson DH (1980) The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. *J Acoust Soc Am* 68:1115–1122.
- Julesz B (1971) *Foundations of Cyclopean Perception*. Chicago, IL: University of Chicago Press.
- Kaernbach C, Bering C (2001) Exploring the temporal mechanism involved in the pitch of unresolved harmonics. *J Acoust Soc Am* 110:1039–1048.
- Kaernbach C, Demany L (1998) Psychophysical evidence against the autocorrelation theory of auditory temporal processing. *J Acoust Soc Am* 104:2298–2306.
- Kim DO, Molnar CE, Matthews JW (1980) Cochlear mechanics: nonlinear behaviour in two-tone responses as reflected in cochlear-nerve-fibre responses and in ear-canal sound pressure. *J Acoust Soc Am* 67:1704–1721.
- Klein MA, Hartmann WM (1981) Binaural edge pitch. *J Acoust Soc Am* 70:51–61.
- Kohlrausch A, Sander A (1995) Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets. *J Acoust Soc Am* 97:1817–1829.
- Krumbholz K, Patterson RD, Pressnitzer D (2000) The lower limit of pitch as determined by rate discrimination. *J Acoust Soc Am* 108:1170–1180.
- Licklider JCR (1951) A duplex theory of pitch perception. *Experientia* 7:128–133.
- Licklider JCR (1956) Auditory frequency analysis. In: Cherry C (ed), *Information Theory*. New York: Academic Press, pp. 253–268.

- Lin JY, Hartmann WM (1998) The pitch of a mistuned harmonic: evidence for a template model. *J Acoust Soc Am* 103:2608–2617.
- Loeb GE, White MW, Merzenich MM (1983) Spatial cross correlation: a proposed mechanism for acoustic pitch perception. *Biol Cybernet* 47:149–163.
- McFadden D (1986) The curious half octave shift: evidence for a basalward migration of the travelling-wave envelope with increasing intensity. In: Salvi RJ, Henderson D, Hamernik RP, Colletti V (eds), *Basic and Applied Aspects of Noise-Induced Hearing Loss*. New York: Plenum Press, pp. 295–312.
- Meddis R, Hewitt M (1991) Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification. *J Acoust Soc Am* 89:2866–2882.
- Meddis R, O'Mard L (1997) A unitary model of pitch perception. *J Acoust Soc Am* 102:1811–1820.
- Micheyl C, Oxenham AJ (2004) Sequential F0 comparisons between resolved and unresolved harmonics: no evidence for translation noise between two pitch mechanisms. *J Acoust Soc Am*: 116:3038–3050.
- Moore BCJ (1973) Frequency difference limens for short-duration tones. *J Acoust Soc Am* 54:610–619.
- Moore BCJ (1982) *An Introduction to the Psychology of Hearing*. 2nd ed. London: Academic Press.
- Moore BCJ (2003) *An Introduction to the Psychology of Hearing*. 5th ed. London: Academic Press.
- Moore BCJ, Glasberg BR (1988) Effects of the relative phase of the components on the pitch discrimination of complex tones by subjects with unilateral and bilateral cochlear impairments. In: Duifhuis H, Wit H, Horst J (eds), *Basic Issues in Hearing*. London: Academic Press, pp. 421–430.
- Moore BCJ, Glasberg BR (1989) Mechanisms underlying the frequency discrimination of pulsed tones and the detection of frequency modulation. *J Acoust Soc Am* 86: 1722–1732.
- Moore BCJ, Glasberg BR (1990) Frequency discrimination of complex tones with overlapping and non-overlapping harmonics. *J Acoust Soc Am* 87:2163–2177.
- Moore BCJ, Moore GA (2003) Perception of the low pitch of frequency-shifted complexes. *J Acoust Soc Am* 113:977–985.
- Moore BCJ, Ohgushi K (1993) Audibility of partials in inharmonic complex tones. *J Acoust Soc Am* 93:452–461.
- Moore BCJ, Rosen SM (1979) Tune recognition with reduced pitch and interval information. *Q J Exp Psychol* 31:229–240.
- Moore BCJ, Sek A (1994) Effects of carrier frequency and background noise on the detection of mixed modulation. *J Acoust Soc Am* 96:741–751.
- Moore BCJ, Sek A (1996) Detection of frequency modulation at low modulation rates: evidence for a mechanism based on phase locking. *J Acoust Soc Am* 100:2320–2331.
- Moore BCJ, Glasberg BR, Shailer MJ (1984) Frequency and intensity difference limens for harmonics within complex tones. *J Acoust Soc Am* 75:550–561.
- Moore BCJ, Glasberg BR, Peters RW (1985) Relative dominance of individual partials in determining the pitch of complex tones. *J Acoust Soc Am* 77:1853–1860.
- Nabelek IV (1996) Pitch of a sequence of two short tones and the critical pause duration. *Acustica* 82:531–539.
- Ohm GS (1843) Über die Definition des Tones, nebst daran geknüpfter Theorie der Sirene und ähnlicher tonbildender Vorrichtungen. *Ann Phys Chem* 59:513–565.

- Oxenham AJ, Plack CJ (1997) A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing. *J Acoust Soc Am* 101:3666–3675.
- Oxenham AJ, Bernstein JGW, Penagos H (2004) Correct tonotopic representation is necessary for complex pitch perception. *Proc Natl Acad Sci USA* 101:1421–1425.
- Palmer AR, Russell IJ (1986) Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells. *Hear Res* 24:1–15.
- Patterson RD, Wightman FL (1976) Residue pitch as a function of component spacing. *J Acoust Soc Am* 59:1450–1459.
- Patterson RD, Handel S, Yost WA, Datta AJ (1996) The relative strength of the tone and noise components in iterated rippled noise. *J Acoust Soc Am* 100:3286–3294.
- Peters RW, Moore BCJ, Glasberg BR (1983) Pitch of components of complex tones. *J Acoust Soc Am* 73:924–929.
- Plack CJ, Carlyon RP (1995) Differences in frequency modulation detection and fundamental frequency discrimination between complex tones consisting of resolved and unresolved harmonics. *J Acoust Soc Am* 98:1355–1364.
- Plack CJ, White LJ (2000a) Pitch matches between unresolved complex tones differing by a single interpulse interval. *J Acoust Soc Am* 108:696–705.
- Plack CJ, White LJ (2000b) Perceived continuity and pitch perception. *J Acoust Soc Am* 108:1162–1169.
- Plomp R (1964) The ear as a frequency analyzer. *J Acoust Soc Am* 36:1628–1636.
- Plomp R (1967) Pitch of complex tones. *J Acoust Soc Am* 41:1526–1533.
- Plomp R, Mimpfen AM (1968) The ear as a frequency analyzer II. *J Acoust Soc Am* 43:764–767.
- Pollack I (1969) Periodicity pitch for white noise—fact or artifact? *J Acoust Soc Am* 45:237–238.
- Pressnitzer D, Patterson RD (2001) Distortion products and the pitch of harmonic complex tones. In: Breebaart DJ, Houtsma AJM, Kohlrausch A, Prijs VF, Schoonhoven R (eds), *Physiological and Psychophysical Bases of Auditory Function*. Maastricht: Shaker, pp. 97–104.
- Pressnitzer D, Patterson RD, Krumholz K (2001) The lower limit of melodic pitch. *J Acoust Soc Am* 109:2074–2084.
- Pressnitzer D, de Cheveigné A, Winter IM (2002) Perceptual pitch shifts for sounds with similar waveform autocorrelation. *Acoust Res Lett Online* 3:1–6.
- Richards VM, Zhu S (1994) Relative estimates of combination weights, decision criteria, and internal noise based on correlation coefficients. *J Acoust Soc Am* 95:423–434.
- Ritsma RJ (1962) Existence region of the tonal residue. I. *J Acoust Soc Am* 34:1224–1229.
- Ritsma RJ (1963) Existence region of the tonal residue. II. *J Acoust Soc Am* 35:1241–1245.
- Ritsma RJ (1967) Frequencies dominant in the perception of the pitch of complex sounds. *J Acoust Soc Am* 42:191–198.
- Robles L, Ruggero MA, Rich NC (1997) Two-tone distortion on the basilar membrane of the chinchilla cochlea. *J Neurophysiol* 77:2385–2399.
- Rossing TD, Houtsma AJM (1986) Effects of signal envelope on the pitch of short sinusoidal tones. *J Acoust Soc Am* 79:1926–1933.
- Ruggero MA, Rich NC, Recio A, Narayan SS, Robles L (1997) Basilar-membrane responses to tones at the base of the chinchilla cochlea. *J Acoust Soc Am* 101:2151–2163.

- Schouten JF (1938) The perception of subjective tones. *Proc Kon Akad Wetenschap* 41: 1086–1093.
- Schouten JF (1940) The residue and the mechanism of hearing. *Proc Kon Akad Wetenschap* 43:991–999.
- Schouten JF (1970) The residue revisited. In: Plomp R, Smoorenburg GF (eds), *Frequency Analysis and Periodicity Detection in Hearing*. Leiden, The Netherlands: Sijthoff, pp. 41–54.
- Schouten JF, Ritsma RJ, Cardozo BL (1962) Pitch of the residue. *J Acoust Soc Am* 34: 1418–1424.
- Schroeder MR (1970) Synthesis of low peak-factor signals and binary sequences with low autocorrelation. *IEEE Trans Inform Theory* 16:85–89.
- Seebeck A (1841) Beobachtungen über einige bedingungen der entstehung von tönen. *Ann Phys Chem* 53:417–436.
- Sek A, Moore BCJ (1995) Frequency discrimination as a function of frequency, measured in several ways. *J Acoust Soc Am* 97:2479–2486.
- Shackleton TM, Carlyon RP (1994) The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. *J Acoust Soc Am* 95:3529–3540.
- Shamma SA (1985a) Speech processing in the auditory system. I: The representation of speech sounds in the responses in the auditory nerve. *J Acoust Soc Am* 78:1612–1621.
- Shamma SA (1985b) Speech processing in the auditory system. II: Lateral inhibition and the central processing of speech evoked activity in the auditory nerve. *J Acoust Soc Am* 78:1622–1632.
- Shamma S, Klein D (2000) The case of the missing pitch templates: how harmonic templates emerge in the early auditory system. *J Acoust Soc Am* 107:2631–2644.
- Siegel RJ (1965) A replication of the mel scale of pitch. *Am J Psychol* 78:615–620.
- Smoorenburg GF (1970) Pitch perception of two-frequency stimuli. *J Acoust Soc Am* 48:924–941.
- Stevens SS (1935) The relation of pitch to intensity. *J Acoust Soc Am* 6:150–154.
- Stevens SS, Volkman J, Newman EB (1937) A scale for the measurement of the psychological magnitude of pitch. *J Acoust Soc Am* 8:185–190.
- Terhardt E (1971) Pitch shifts of harmonics, an explanation of the octave enlargement phenomenon. *Proc 7th ICA, Budapest, Hungary*, 621–624.
- Terhardt E (1974) Pitch, consonance, and harmony. *J Acoust Soc Am* 55:1061–1069.
- Terhardt E (1979) Calculating virtual pitch. *Hear Res* 1:155–182.
- Terhardt E, Fastl H (1971) Zum Einfluss von Störtönen und Störgeräuschen auf die Tonhöhe von Sinustönen. *Acustica* 25:53–61.
- Terhardt E, Stoll G, Seewann M (1982a) Pitch of complex signals according to virtual pitch theory. *J Acoust Soc Am* 71:671–678.
- Terhardt E, Stoll G, Seewann M (1982b) Algorithm for extraction of pitch salience from complex tonal signals. *J Acoust Soc Am* 71:679–688.
- van de Par S, Kohlrausch A (1997) A new approach to comparing binaural masking level differences at low and high frequencies. *J Acoust Soc Am* 101:1671–1680.
- Verschuure J, van Meeteren AA (1975) The effect of intensity on pitch. *Acustica* 32: 33–44.
- Viemeister NF (1979) Temporal modulation transfer functions based upon modulation thresholds. *J Acoust Soc Am* 66:1364–1380.
- Viemeister NF, Wakefield GH (1991) Temporal integration and multiple looks. *J Acoust Soc Am* 90:858–865.

- Ward WD (1954) Subjective musical pitch. *J Acoust Soc Am* 26:369–380.
- White LJ, Plack CJ (1998) Temporal processing of the pitch of complex tones. *J Acoust Soc Am* 103:2051–2063.
- Wiegand L (2001) Searching for the time constant of neural pitch extraction. *J Acoust Soc Am* 109:1082–1091.
- Wier CC, Jesteadt W, Green DM (1977) Frequency discrimination as a function of frequency and sensation level. *J Acoust Soc Am* 61:178–184.
- Yates GK, Winter IM, Robertson D (1990) Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range. *Hear Res* 45:203–220.
- Yost WA, Patterson RD, Sheft S (1996) A time-domain description for the pitch strength of iterated rippled noise. *J Acoust Soc Am* 99:1066–1078.
- Yost WA, Patterson R, Sheft S (1998) The role of the envelope in processing iterated rippled noise. *J Acoust Soc Am* 104:2349–2361.
- Zwicker E (1970) Masking and psychological excitation as consequences of the ear's frequency analysis. In: Plomp R, Smoorenburg GF (eds), *Frequency Analysis and Periodicity Detection in Hearing*. Leiden: Sijthoff, pp. 376–394.
- Zwicker E, Fastl H (1990) *Psychoacoustics—Facts and Models*. Berlin: Springer-Verlag.

# Comparative Aspects of Pitch Perception

WILLIAM P. SHOFNER

## 1. Introduction

It should be evident from a glance at the topics covered in this volume regarding human pitch perception (Plack and Oxenham, Chapter 2; Moore and Carlyon, Chapter 7; Darwin, Chapter 8; Bigand and Tillmann, Chapter 9) that pitch perception is an umbrella covering a broad range of perceptual attributes. Do animals possess pitch perceptions similar to those of human listeners? The use of the word “pitch” in conjunction with the word “animals” is somewhat anthropomorphic. Indeed, Fay (1995) has argued that placing a label on the animal perception that is analogous in some manner with the label used to describe human perception is not particularly informative. The perception of complex, periodic sounds by animals may or may not be similar to human pitch perception. A more appropriate question to address in animals is the following: “Are the stimulus features that influence perception in human listeners the same features that influence perception in animals?” In other words, what stimulus features control the behavioral response of the animal, and how does the behavioral response change as these features change systematically? Comparing and contrasting the stimulus features that influence animal discriminations and perceptions with those that influence human discriminations and perceptions can then give us insights into the similarities and differences in the mechanisms underlying the perceptual dimensions. This chapter provides an overview of the studies in vertebrate animals as they relate to some of these perceptual attributes of pitch. The purpose of this chapter is not to review the animal behavioral data in order to provide the animal perception with a “pitch” label, but rather to present the behavioral data in order to answer the types of questions raised above. For conciseness, pitch (i.e., no quotes) will be used when referring to the human perception, and ‘pitch’ (i.e., single quotes) will be used when referring to the animal perception.

In Volume 4 of the *Springer Handbook of Auditory Research*, Fay (1994a) described two goals of psychophysical studies in animals. One goal is to develop an appropriate “animal model” for human hearing, in order to then use



the animal model to study the neurophysiological basis of human hearing. Understanding behavior in animals is a necessary and important conceptual bridge between behavioral studies in human listeners and neurophysiological experiments in animals. The neurophysiological responses of auditory neurons to stimulus features that are important for pitch perception are discussed in this volume (see Winter, Chapter 4) and therefore are not presented in this chapter. The second goal of animal psychophysics, referred to by Fay as “comparative hearing research,” is to study hearing in animals in order to understand hearing as a “general biological phenomenon” (Fay 1994a). It is the comparative hearing approach that is emphasized in this chapter, and any references to human pitch perception are made in an effort to place the appropriate human data within the larger context of the animal data. That is, in this chapter, humans are viewed as simply another mammalian species. Some of the animal behavioral studies discussed have been carried out specifically with pitch-related issues in mind, whereas others have not, but are relevant to this overview given the nature of the periodic stimuli used. In an effort to facilitate comparisons across phylogeny and provide a more integrated discussion, the approach of this chapter is to present the research related to each specific perceptual attribute for all vertebrates studied, rather than describing all of the pitch-related research for each individual vertebrate class separately.

## 2. Methodology

### *2.1 Training and Conditioning*

In any psychophysical experiment, it is important that the response of the subject is controlled by some physical dimension of the stimulus. In human psychophysical experiments, the experimenter often discusses with the subject the nature of the experiment and the stimulus features that are important. In other words, the subject is informed verbally as to what stimulus cues should be attended to during testing. In animal psychophysical experiments, it is just as important that the behavioral response of the animal be under stimulus control, but verbal communication regarding the specific stimulus features is not available to the experimenter. The animal must be trained or conditioned to give the appropriate behavioral response to the particular stimulus dimension that will be varied in the experiment. Training or conditioning of the behavioral response generally falls into three categories.

In classical conditioning (also called Pavlovian conditioning), an unconditioned stimulus evokes a natural or reflexive response from the animal. This response is called the unconditioned response, since it occurs without any training or conditioning. The presentation of the unconditioned stimulus is then paired with the presentation of an experimental stimulus, known as the conditioned stimulus. Over time, the conditioned stimulus will evoke a response similar to the unconditioned response in the absence of the unconditioned stim-

ulus; this response is referred to as the conditioned response. For example, in the behavioral work by Fay and his colleagues (see below), goldfish show a suppression of respiration (unconditioned response) when presented with a mild electric shock (unconditioned stimulus). When the shock is paired with a tone, the goldfish will show respiratory suppression (conditioned response) when the tone is presented without an unconditioned stimulus.

In instrumental avoidance conditioning, the animal is trained to avoid the shock by making some behavioral response. For example, an animal is placed in a cage and must shuttle back and forth across the cage in order to avoid the shock. In this example, the electric shock is preceded by the presentation of the conditioned stimulus (i.e., tone), and the animal learns to avoid the shock when it hears the tone by crossing to the other side of the cage. This avoidance procedure has also been used on animals that are restricted in the amount of water they receive *ad libitum*. During training, they are given access to a water spout and can drink freely; in this case, the animals learn to cease drinking when the conditioned stimulus is presented in order to avoid the electric shock.

In operant conditioning, the amount of food (or water) the animal receives *ad libitum* is generally restricted, and then food (or water) is used as a reward during training and testing. Animals are trained to make an overt response (i.e., key press, release of a lever, pecking a disk) in order to receive the reward. The reward is paired with the appropriate stimulus and the animal receives the reward when the operant response is made during the presentation of the stimulus.

Classical conditioning and avoidance conditioning procedures both have the advantage that animals learn the behavioral task relatively quickly, but have the disadvantage that over long periods of time the behavior often breaks down so that the animal is no longer under stimulus control. Operant conditioning has the advantage that the behavior does not break down over prolonged periods of time, but has the disadvantage that training generally requires long periods of time in order to establish the behavioral response. Psychophysical procedures that are commonly used in human experiments, such as the two-interval forced-choice procedure, have been difficult to adapt in animal behavioral paradigms. Most animal psychophysical experiments are based on a Go/No Go paradigm; in these procedures, the animal must wait for some random period of time after which the signal may or may not be presented. If the animal perceives the signal, it makes the appropriate operant response (i.e., Go); if the animal does not perceive the signal, it does not make the response (i.e., No Go).

## *2.2 Distinction Between Discrimination and Stimulus Generalization*

The animal psychophysical studies discussed in this chapter typically fall into two general categories, namely discrimination studies and stimulus generalization studies. Discrimination studies are generally concerned with measuring

acuity to changes in the signal; that is, the experimenter is interested in estimating when the animal can just detect a change in the signal along some physical dimension of the stimulus. In discrimination experiments, animals always receive feedback when they make a correct behavioral response (e.g., feedback can be obtaining a food reward or successfully avoiding an electric shock). Thresholds are defined for some criterion level of response, and these thresholds can often be compared directly to those obtained in human psychophysical experiments. Because there are correct and incorrect responses that can be made by the animal, these experiments are objective in nature. Comparisons between animal and human discrimination data are always plagued by questions regarding procedural differences between animal and human paradigms. As a procedural control, some animal behavioral experiments also collect data from human listeners using the animal behavioral task. The human data obtained using animal procedures are more often than not similar to the data obtained in traditional human experimental paradigms and serve to validate the data obtained from the animal psychophysical procedures.

Pitch is a percept, and as such, studies addressing questions concerning pitch in human listeners have often used subjective procedures such as pitch matching and scaling methods (e.g., magnitude estimation). These types of procedures do not have a direct counterpart in animal studies the way that objective discrimination experiments do. However, perceptual questions in animals can be addressed using stimulus generalization paradigms. In stimulus generalization paradigms, animals are trained to respond to a specific stimulus, and then responses are measured to probe or test stimuli that vary systematically along one or more stimulus dimensions (Malott and Malott 1970). A systematic change in behavioral response along the physical dimension of the stimulus is known as a generalization gradient and is consistent with the hypothesis that the animal possesses a perceptual dimension related to the physical dimension of the stimulus (Guttman 1963). A generalization gradient is often interpreted to indicate similarities in an animal's perception between probe and training stimuli. Probe stimuli that evoke similar behavioral responses as the training stimulus indicate a perceptual equivalence or perceptual invariance (see Hulse 1995) among these stimuli. In other words, stimuli that are perceptually invariant or equivalent contain a stimulus feature that is perceived to be functionally equal among the stimuli (Hulse 1995). Thus, data from stimulus generalization paradigms can give insights into what features of the stimulus are being attended to or analyzed during testing and can be used to indicate what stimulus features control the behavioral response of the animal. It should be noted that unlike discrimination experiments in which animals receive feedback (e.g., food reward, no electric shock) for correct behavioral responses, responses to probe stimuli in generalization experiments are not rewarded, because they are considered to be neither correct nor incorrect (i.e., they are subjective responses).

### 3. Periodicity Discrimination and Perception in Vertebrate Animals

Many naturally occurring sounds are produced by vibratory objects that generate complex acoustic waveforms having some form of temporal periodicity, and an important perception closely related to periodicity is, of course, pitch. This section provides an overview of the studies in vertebrate animals as they relate to some of the perceptual attributes evoked by periodic sounds.

#### 3.1 Frequency Perception of Single Tones

Many reviews on pitch perception in human listeners begin with a discussion of frequency discrimination of single tones. Single-tone frequency discrimination has been reviewed in the *Springer Handbook of Auditory Research* series for mammals (Volume 4), fish and anurans (Volume 11), and birds and reptiles (Volume 13), and therefore will not be extensively re-reviewed in this chapter. In general, the data show that animals can discriminate between the frequencies of single tones, but that their thresholds for discrimination are often higher than those in human listeners (see Fig. 3.1). Perhaps a more interesting starting point for questions regarding ‘pitch’ perception in animals is not whether animals can discriminate frequencies, but whether animals possess a perceptual dimension related to tone frequency, similar to spectral pitch for single tones in human listeners.

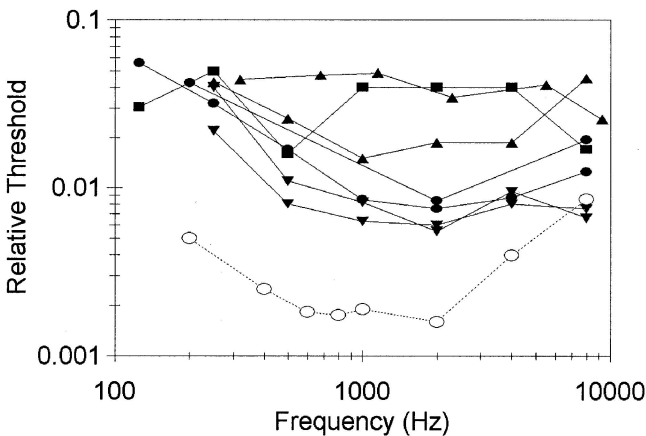


FIGURE 3.1. Frequency discrimination thresholds for tones among common laboratory mammals generally considered to have good low-frequency hearing abilities. Threshold is expressed as relative threshold, which is the fractional change in frequency (i.e.,  $\Delta f/f$ , where  $\Delta f$  is the difference limen). Filled squares show guinea pig data; filled circles show cat data; filled triangles show chinchilla data; filled inverted triangles show monkey data. Open circles show human data. Data were compiled from Fay (1988).

Single-tone frequency perception has been studied for animals in stimulus generalization experiments. In these experiments, animals are trained to respond to a single tone of a specific frequency and then tested in a stimulus generalization paradigm using single tones of varying frequencies. If animals possess a perceptual dimension related to the physical dimension of frequency, then stimulus generalization gradients would be expected to show large behavioral responses to the tone at the training frequency with monotonic decreases in response as the frequency of the test tone changes from the training tone. Generalization gradients such as these have been obtained for rats (*Rattus norvegicus*, Blackwell and Schlosberg 1943), pigeons (*Columbia livia*, Jenkins and Harrison 1960), starlings (*Sturnus vulgaris*, Cynx 1993), and goldfish (*Carassius auratus*, Fay 1992a). Thus, perceptual dimensions related to tone frequency appear to exist across vertebrates. Fay et al. (1996) have also shown that when goldfish are conditioned using a 400-Hz single tone and then tested with ramped and damped sinusoids of the same frequency, they generalize more to ramped sinusoids than to damped sinusoids. A damped sinusoid has an envelope with a sudden onset followed by an exponential decay of a specified time; a ramped sinusoid has an envelope with an exponential rise of a specified time followed by a sudden offset. The generalization gradients obtained by Fay et al. (1996) suggest that goldfish perceive ramped sinusoids to be more tonal than damped sinusoids, a finding similar to that observed in human listeners (Patterson 1994a).

A subjective pitch scale has been derived for human listeners using single tones (Stevens and Volkmann 1940; see also Plack and Oxenham, Chapter 2). Dooling et al. (1987a) have derived a “pitch” scale for budgerigars (*Melopsittacus undulatus*) by applying a multidimensional scaling technique to operant behavioral methods. Figure 3.2A compares the derived ‘pitch’ scales for budgerigars for a narrow range of single-tone frequencies from 2000 to 4000 Hz with a broader range of frequencies from 1000 to 5700 Hz. The frequency range between 2000 and 4000 Hz is the spectral region where budgerigars are most sensitive and show the greatest frequency selectivity (see Dooling et al. 1987a). Over this narrow frequency range, the ‘pitch’ scale is relatively linear on a linear-log coordinate system, whereas the ‘pitch’ scale derived for the broader range of frequencies can be described by three separate linear functions. The overall shape of the ‘pitch’ scale for the broad frequency range might also be described with a sigmoidally shaped function rather than three separate linear functions. In that respect, it is interesting to note that on a linear-log coordinate system, the mel pitch scale for human listeners also has a sigmoidal shape over a wide range of frequencies (see Stevens and Volkmann 1940). Figure 3.2B compares the budgerigar ‘pitch’ scale to the pitch scale obtained from human listeners for frequencies from 2000 to 4000 Hz using the same multidimensional scaling procedure. Over this narrow frequency range, both the budgerigar and human scales can be accounted for by a linear function, but the slope of the budgerigar function is significantly steeper than the slope for the human function (Dooling et al. 1987a). Dooling et al. (1987a) conclude that a change in single-

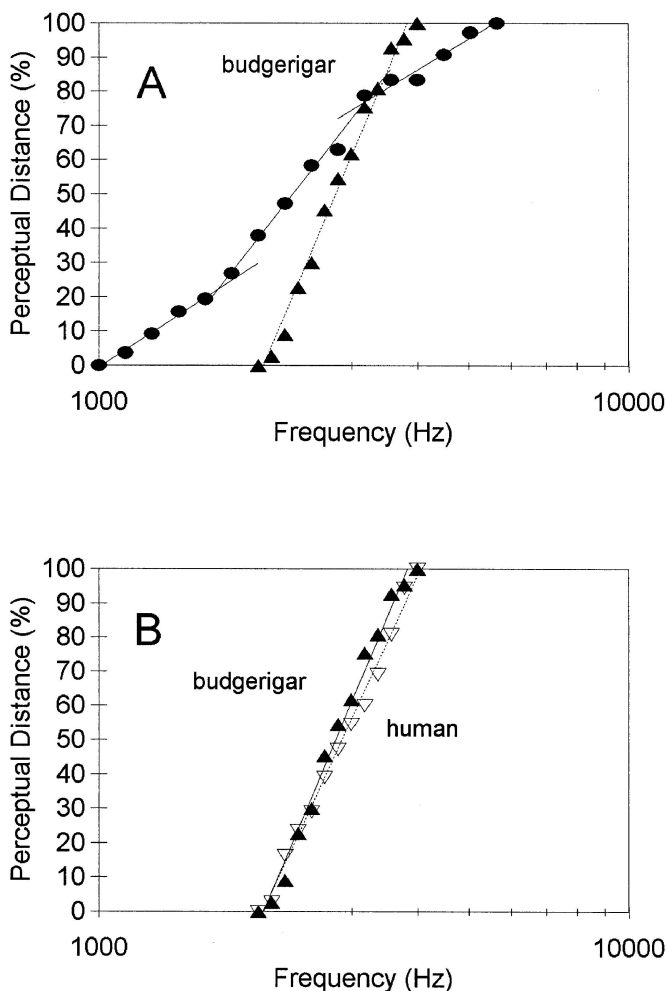


FIGURE 3.2. 'Pitch' scales derived using a multidimensional scaling procedure by Dooling et al. (1987a) for budgerigars and human listeners. Percent perceptual distance is derived from a multidimensional scaling analysis. Data are plotted on a linear-log axis which has been used to describe the mel scale in human listeners (see Stevens and Volkman 1940). (A) 'Pitch' scales obtained for budgerigars. *Filled triangles* and *dotted regression line* show data obtained for 13 single tones between 2000 and 4000 Hz; tone frequencies change in 1/12-octave steps. *Filled circles* and *solid regression lines* show data obtained for 16 single tones between 1000 and 5700 Hz; tone frequencies change in 1/6 octave steps. (B) Comparison of 'pitch' scales obtained from budgerigars and human listeners for 13 single tones between 2000 and 4000 Hz; tone frequencies change in 1/12-octave steps. *Filled triangles* and *solid regression line* show data from budgerigars; *open inverted triangles* and *dotted regression line* show data from human listeners. The regression line through the budgerigar data is  $y = 368x - 1218$  ( $r^2 = 0.988$ ); the regression line through the human data is  $y = 335x - 1110$  ( $r^2 = 0.993$ ). Modified from Figures 4 and 5 of Dooling et al. (1987a) with the authors' permission. © 1987 by the American Psychological Association. Adapted with permission.

tone frequency gives rise to a more salient 'pitch' change in budgerigars than in human listeners.

### 3.2 *Discrimination of Fundamental or Modulation Frequency of Complex, Periodic Sounds*

A harmonic tone complex comprised of a fundamental frequency (F0) and successive higher harmonics evokes the perception of a single sound source in human listeners in which the pitch of the sound is matched to the F0. The effect of F0 on behavior has also been studied in anuran amphibians. Using the evoked calling response of the bullfrog (*Rana catesbeiana*) to synthetic mating calls, Capranica (1966) showed that the largest behavioral response was obtained with mating calls having a F0 of 100 Hz, which corresponds to the waveform periodicity of the natural call. Synthetic mating calls having other F0s between 25 and 200 Hz were less effective in evoking the vocal response. In contrast to these results, Gerhardt (1981) found no difference in behavioral responses of barking treefrogs (*Hyla gratiosa*) based on phonotaxis to synthetic mating calls having F0s of 500 Hz or 250 Hz.

Frequency discrimination of harmonic complex tones has been studied in chinchillas (*Chinchilla laniger*, Shofner 2000). Chinchillas were trained to discriminate a 250-Hz F0 tone complex from a tone complex having a higher F0. The tone complexes were comprised of the F0 and the 2nd through 10th harmonics with individual components added in cosine-starting phase. Psychometric functions were obtained at different overall sound pressure levels, and discrimination thresholds were independent of overall level. Psychometric functions were also obtained for frequency discrimination of a single 250-Hz tone. Estimates of frequency difference limens from the psychometric functions indicate that thresholds in chinchillas were lower for harmonic tone complexes than for a single tone at the F0. This finding is similar to results described previously for human listeners (Flanagan and Saslow 1958; Henning and Grosberg 1968; Fastl and Weinberger 1981; Moore et al. 1984; Spiegel and Watson 1984). Presumably, in human listeners, the information from each of the individual harmonic components is integrated to form a single source having a pitch, which produces better frequency discrimination (see Moore et al. 1984; Moore 1993). The similarity in the results suggests that the neural mechanisms for frequency discrimination of complex tones are not different between chinchilla and human auditory systems.

One class of stimuli that has been used to study periodicity discrimination in animals is sinusoidally amplitude-modulated (SAM) sound. Modulation detection of SAM sounds has been studied across many vertebrate species, but these studies are probably related more to intensity processing than pitch perception and, therefore, are not discussed here. More relevant to pitch perception are studies of the discrimination of modulation frequencies of SAM sounds. Schulze and Scheich (1999) studied modulation discrimination using SAM tones in the gerbil (*Meriones unguiculatus*). Gerbils were trained to discriminate be-

tween two SAM tones having a carrier frequency of 2 kHz, but differing in modulation frequencies. Discrimination performance was high when modulation frequencies differed by one octave. It was observed that gerbils learned the discrimination faster when the modulation frequencies were below 100 Hz, but took longer to reach a high performance level when modulation frequencies were above 100 Hz.

SAM noise is generated when a wideband noise is modulated by a single tone, and this type of stimulus can evoke the perception of pitch in human listeners (Burns and Viemeister 1976, 1981). Periodicity information exists only for the modulation frequency found in the stimulus envelope; there are no long-term spectral cues for the modulation frequency. This type of frequency discrimination is often referred to as rate discrimination, and Figure 3.3 summarizes the rate discrimination thresholds across vertebrates studied (macaque monkey [*Macaca*], Moody 1994; chinchilla, Long and Clark 1984; goldfish, Fay and Passow 1982, Fay 1982; budgerigar, Dooling and Searcy 1981). In general, average rate discrimination thresholds for vertebrate animals fall above those of human listeners, although the function for the budgerigar appears to fall within the range of human thresholds. Monkey thresholds for modulation frequencies around 80 to 100 Hz also appear to fall within the range of human thresholds.

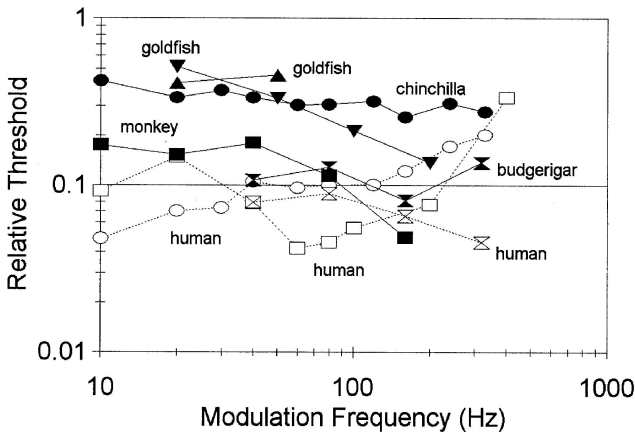


FIGURE 3.3. Rate discrimination thresholds for SAM noise among vertebrates. Threshold is expressed as relative threshold, which is the fractional change of the modulation frequency (i.e.,  $\Delta f_{\text{modulation}} / f_{\text{modulation}}$ ). *Filled squares* show monkey data from Moody (1994); *filled circles* show chinchilla data from Long and Clark (1984); *filled hourglasses* show budgerigar data from Dooling and Searcy (1981); *filled triangles* show goldfish data from Fay (1982). *Open squares* show human data from Formby (1985); *open circles* show human data from Long and Clark (1984); *open hourglasses* shows human data from Dooling and Searcy (1981). The *filled inverted triangles* show goldfish data from Fay and Passow (1982) obtained for filtered Gaussian noise presented at repetition rates corresponding to the modulation frequency.



Over the range of modulation frequencies studied, the Weber fraction (i.e., relative threshold) appears to be relatively constant for budgerigars and chinchillas, but not for monkeys and goldfish. Also note that over a similar frequency range of 100 to 200 Hz, there is about one order of magnitude difference between the Weber fractions for rate discrimination (Fig. 3.3) and those for single-tone frequency discrimination (Fig. 3.1) for all species.

### *3.3 Perception of the Missing Fundamental*

One noteworthy attribute of human pitch perception is known as the pitch of the missing fundamental. In human listeners, when a harmonic tone complex contains no acoustic energy at the F0, the perceived pitch of the tone complex is still matched to the corresponding missing F0 (see Plack and Oxenham, Chapter 2). The following section describes psychophysical experiments in animals using stimuli that are known to evoke the perception of the missing fundamental in human listeners.

One type of complex sound having a missing fundamental is a SAM tone. The SAM tone constitutes a three-component harmonic tone complex with a missing fundamental at the modulation frequency. Fay (1972) studied the perception of SAM tones in goldfish using a stimulus generalization paradigm. Goldfish were trained to respond to a SAM tone having a depth of modulation of 100%, and the carrier frequency was fixed at either 400 Hz or 1000 Hz. Goldfish showed a large behavioral response to the training SAM tone, but showed a decrease in behavioral response when tested with SAM tones having the same carrier frequency, but varying in modulation frequency. Goldfish were also trained to respond to a 40-Hz single tone, and then were tested using 100% SAM tones having a fixed carrier frequency of 1000 Hz. Goldfish showed a large behavioral response to 40-Hz modulated SAM tones, but showed a decrease in behavioral response when presented with SAM tones having other modulation frequencies between 15 and 80 Hz. Although these results do not demonstrate a perception of the missing fundamental in goldfish, they do suggest that goldfish possess a perceptual dimension along the physical dimension of envelope periodicity. That is, changing the F0 alters the perception in goldfish.

Fay (1995) trained two groups of goldfish to respond to a 100-Hz single tone. One group was tested in a stimulus generalization paradigm with tones of other frequencies. These goldfish showed a decrease in behavioral response as tone frequency increased away from 100 Hz, suggesting that goldfish possess a perceptual dimension along the physical dimension of frequency. The second group was tested in the generalization paradigm on five different harmonic complex tones having a F0 of 100 Hz. One complex contained the F0, whereas the other tone complexes had a missing fundamental. For these goldfish, the behavioral responses to all of the tone complexes were weak; that is, goldfish trained on a 100-Hz single tone did not generalize to complex tones having F0s of 100 Hz. The failure to generalize to the complex tones does not necessarily imply that goldfish do not perceive the missing fundamental, but it does suggest that timbre-

like cues of spectral location may be more salient. It is interesting to note that starlings can discriminate between complex tones comprised of the fundamental and varying harmonic components (Braaten and Hulse 1991), suggesting that spectral location (i.e., 'timbre') is also a salient cue in birds. When goldfish are conditioned to respond to a periodic pulse train at a given repetition rate, they showed large responses to pulse trains at the conditioning repetition rate, and monotonically decreasing responses as repetition rate varied from the conditioning rate (Fay 1994b). These generalization gradients are consistent with the hypothesis that goldfish possess a perceptual dimension along the physical dimension of pulse repetition rate (i.e., F0).

Cynx and Shapiro (1986) showed that starlings appear to have a missing fundamental percept. Starlings were trained to peck a lighted-disk during the presentation of a 625-Hz complex tone with the missing fundamental and cease pecking during the presentation of a 400-Hz complex tone with a missing fundamental. The harmonic components of the tone complexes were varied; thus, the discrimination could be done using only the perception of the missing fundamental as the cue. Birds were then tested in a generalization paradigm in which single tones at 625 Hz or 400 Hz were presented. Starlings showed no significant difference in their behavioral responses between the 625-Hz tone complex and the 625-Hz single tone, but showed a significant difference in behavioral responses between the 625-Hz tone complex and the 400-Hz single tone. These findings are consistent with the perception of the missing fundamental and pitch constancy.

Heffner and Whitfield (1976) and Whitfield (1980) studied the perception of the missing fundamental in cats (*Felis catus*) using SAM tones. Cats were trained to lick a drinking spout to receive a water reward when two single tones alternated between 400 Hz and 342 Hz and were trained to cease drinking to avoid a mild electric shock when the tones alternated between 400 Hz and 458 Hz. That is, cats were trained to drink when the standard frequency decreased and to stop drinking when the standard frequency increased. Cats were then tested using SAM tones in place of the single tones. Figure 3.4A shows the average behavioral results combined from both Heffner and Whitfield (1976) and Whitfield (1980) when the three frequency components of the tone complex and the frequency of the missing fundamental increased or decreased in the *same* direction. Note that the time the cats spent licking the spout was high when the frequencies decreased ( $-/-$ ), but was low when the frequencies increased ( $+/+$ ). Figure 3.4A also shows the average behavioral results obtained when the frequency of the missing fundamental and the three frequency components increased or decreased in *opposite* directions (e.g., the missing fundamental decreases, but the three frequency components increase). Now it can be observed that the time spent licking the spout was high when the missing fundamental decreased ( $-/+$ ), but was low when the missing fundamental increased in frequency ( $+/-$ ) (Fig. 3.4A). In contrast, the time spent drinking is high when the frequencies of the three components increased ( $-/+$ ), but was low when the frequencies of the three components decreased ( $+/-$ ) (Fig. 3.4A). These be-

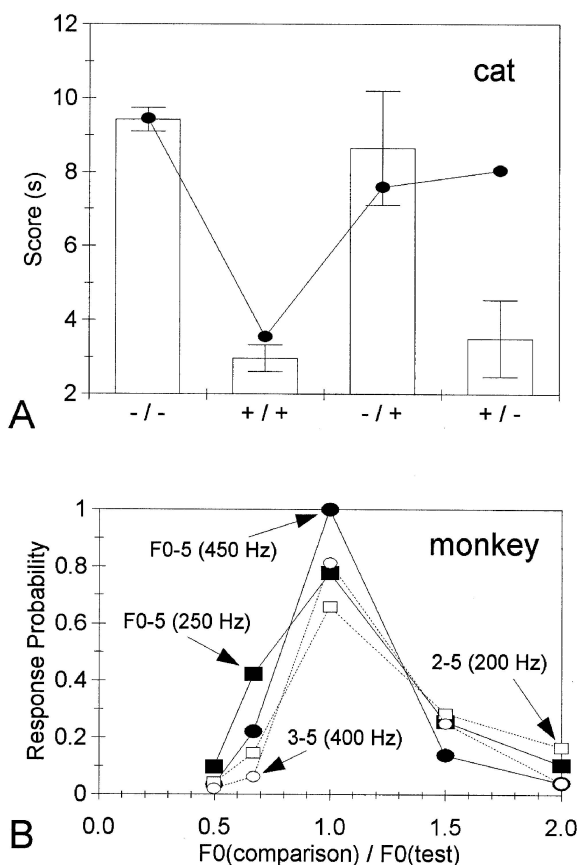


FIGURE 3.4. Behavioral responses illustrating the perception of the missing F0 in mammals. **(A)** Bar graph showing the time spent by cats licking a water spout. Cats were trained to cease licking to avoid a mild electric shock. Scores are averages combined from two cats in Table II of Heffner and Whitfield (1976) and two cats in Table I of Whitfield (1980). Error bars indicate  $\pm 1$  standard deviation. *Filled circles* show the average of two cats after bilateral ablation of the auditory cortex (Whitfield, 1980). The labels on the x-axis indicate the change in frequency for the missing fundamental and the harmonic components of the tone complex. The symbol (-/-) indicates that the missing fundamental and harmonic components both decrease; (+/+) indicates that the missing fundamental and harmonic components both increased; (-/+) indicates that the missing fundamental decreased, but the harmonic components increased; (+/-) indicates that the missing fundamental increased, but the harmonic components decreased. **(B)** Stimulus generalization gradients obtained from monkeys (Tomlinson and Schwarz, 1988). *Filled circles* and *filled squares* show the gradients in behavioral responses obtained when the test stimulus was a harmonic tone complex comprised of the F0 and the 2nd through 5th harmonics for F0s of 450 Hz and 250 Hz, respectively. *Open symbols* show the generalization gradients obtained when the F0 of the test stimulus was missing. The test tone complexes were comprised of the 2nd through 5th harmonics with a 200 Hz missing fundamental (*open circles*) or comprised of the 3rd through 5th harmonics with a 400 Hz missing fundamental. Modified from Figure 2 of Tomlinson and Schwarz (1988) with the authors' permission.

havioral results indicate that the perception of the missing fundamental controlled the behavioral response of the cats, rather than the actual frequencies of the tone complex, because the cats were initially trained to cease drinking (i.e., contact times should be small) when the frequencies increased.

Whitfield (1980) demonstrated that the auditory cortex was important in the perception of the missing fundamental in cats. After bilateral ablation of primary and secondary auditory cortices, cats no longer retained the ability to discriminate the single tones, but were able to relearn the discrimination. These behavioral results are consistent with those obtained by others (Butler et al. 1957; Cranford et al. 1976; Ohm et al. 1999) for single-tone frequency discrimination following bilateral ablation of auditory cortex. Figure 3.4A also shows the behavioral results of the cats following bilateral ablation of the auditory cortex. Similar to normal cats, the time that lesioned cats spent licking the spout was high when the frequencies of both the missing fundamental and harmonics decreased ( $-/-$ ), but was low when the frequencies increased ( $+/+$ ). However, when the frequency of the missing fundamental and the three frequency components changed in *opposite* directions, the cats no longer showed a behavioral response consistent with the missing fundamental. Now it can be observed that the time spent licking the spout was high when the missing fundamental either decreased ( $-/+$ ) or increased ( $+/-$ ). The findings in lesioned cats suggest that the perception of the missing fundamental no longer controlled the behavioral response, but rather the behavior was controlled by the changes in the spectral locations of the three harmonic components of the tone complex. Similar findings have been obtained from human listeners having temporal lobe lesions (Zatorre 1988). Thus, the auditory cortex is important for pitch perception, but may not be essential for frequency discrimination. More recently, Tramo et al. (2002) have shown that frequency discrimination thresholds are elevated in patients with bilateral auditory cortex lesions, but not in patients with unilateral lesions. It is also interesting to note that discrimination performance for frequency-modulated tones is significantly reduced in gerbils following bilateral ablation of the auditory cortex (Ohm et al. 1999). Also, monkeys can discriminate intermittent noise from noninterrupted noise (for rates between 10 and 80 pulses per second), but fail to re-learn the discrimination following bilateral auditory cortex ablation (Symmes 1966).

Tomlinson and Schwarz (1988) presented rhesus monkeys (*Macaca mulata*) with two successive complex tones, and trained the monkeys to push a button after the onset of the second-tone complex if the second-tone complex had the same F0 as the first-tone complex. The first-tone complex was a test stimulus in which the F0 was fixed, but was either present or missing. The second-tone complex was the comparison stimulus in which the F0 varied, but was always present. Figure 3.4B shows the average stimulus generalization gradients obtained. When the F0 of the test equaled that of the comparison tone complex (i.e., ratio is 1) and the F0 was present in the test tone complex, the probability of a behavioral response was the highest. As the difference between the F0s of the comparison and test complexes increased (i.e., as the ratio deviated from 1),

there was a systematic decrease in the behavioral response. More importantly, similar generalization gradients were obtained when the F0 of the test complex was missing (Fig. 3.4B), suggesting a perception of the missing fundamental.

The results described previously in this section are consistent with the hypothesis that animals possess a pitch percept corresponding to the frequency of the missing fundamental. However, as described by Plack and Oxenham in Chapter 2, the pitch of the missing fundamental in human listeners remains even in the presence of low-frequency masking noise. Since none of the above animal studies used low-frequency masking noise, a potential role of combination tones generated by the nonlinearities in the auditory organs cannot be ruled out at present.

### 3.4 *Mistuned Harmonics and Analytic Listening*

As described previously, a harmonic tone complex evokes the perception of a single sound source in human listeners having a pitch matched to the F0. This perception of a single sound source is a form of synthetic listening. However, if the frequency of one of the components is changed such that it no longer is related harmonically to the other components, then this mistuned harmonic can be heard as a separate sound source from the harmonic background (see Plack and Oxenham, Chapter 2). The perceived pitch of the mistuned harmonic is matched to the frequency of the mistuned harmonic and is a form of analytic listening.

The effect of mistuned harmonics has been studied in the mating calls of anuran amphibians. Simmons and Bean (2000) measured the evoked vocal response of the bullfrog to synthetic mating calls having an F0 of 100 Hz. Synthetic calls were comprised of all 22 harmonics of the F0 between 100 and 2200 Hz. The mating call of the bullfrog contains large spectral peaks at 200 Hz and 1400 Hz with a spectral valley around 500 to 600 Hz; these spectral peaks are necessary to evoke the calling response in male bullfrogs. Large evoked calling responses were obtained from male bullfrogs to the synthetic mating call, but synthetic calls with mistuned harmonics evoked weaker vocal responses. Significant decreases in the number of evoked responses were obtained when either the harmonic component at 200 Hz or the component at 1400 was mistuned. These authors concluded that it was likely that bullfrogs detected differences between the harmonic calls and mistuned-harmonic calls through changes in the temporal envelope of the synthetic call.

In a similar study, Simmons (1988) used a reflex modification technique to measure detection thresholds in the green treefrog (*Hyla cinerea*) for two-tone complexes. Each of the two-tone complexes were comprised of frequency components that fell within the range of the two spectral peaks in the mating call. The mating call of the green treefrog has large spectral peaks around 900 Hz and 3000 Hz, and both of these peaks are necessary for the call to evoke behavioral responses. Low detection thresholds were obtained when the tones of the complexes were harmonically related (900 + 3000 Hz and 828 + 2760 Hz).

In contrast, high detection thresholds were obtained if the frequencies were inharmonically related ( $831 + 3100$  Hz).

Mistuned harmonics have also been studied in birds. Lohr and Dooling (1998) compared thresholds for the detection of mistunings in zebra finches (*Taeniopygia guttata*, songbird), budgerigars (nonsongbird), and human listeners. Stimuli were harmonic tone complexes comprised of the first 16 harmonics of either a 570-Hz F0 or a 285-Hz F0. These frequencies were chosen because 570 Hz falls within the range of F0s of the zebra finch call, whereas 285 Hz falls within the range of F0s for human speech. Figure 3.5 summarizes the data for the two F0s. It is clear from this figure that thresholds for detecting mistunings are much higher for human listeners than for either species of birds. The human thresholds are higher than the bird thresholds by factors that range from 3 to 33 (compare human and budgerigar for the 2nd harmonic of the 570-Hz fundamental; compare human and zebra finch for the 7th harmonic of the 570-Hz fundamental). Both species of birds show better performance than human listeners, even when the F0 was chosen to be more characteristic of human speech (i.e., 285 Hz). It should be noted that the human thresholds obtained by Lohr and Dooling (1998) were similar to previously reported thresholds (Moore et al. 1985).

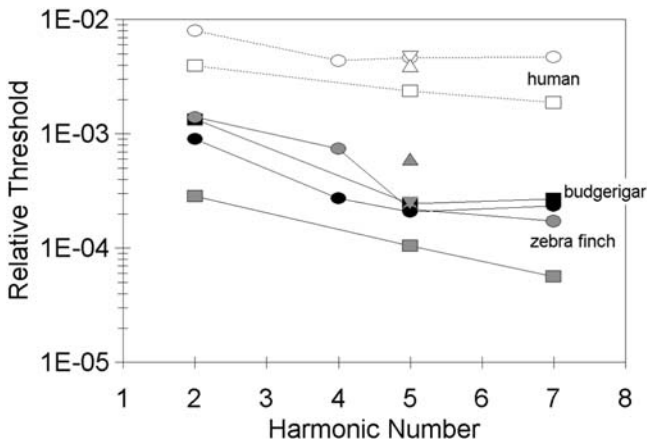


FIGURE 3.5. Detection thresholds for mistuning harmonics in birds and human listeners. Threshold is expressed as relative threshold, which is the fractional change of the harmonic frequency (i.e.,  $\Delta f_{\text{harmonic}} / f_{\text{harmonic}}$ ). Open symbols show human thresholds; gray symbols show zebra finch thresholds; black symbols show budgerigar thresholds. Squares show functions for a 570-Hz F0; circles show functions for a 285-Hz F0. Mistuned harmonics occurred for harmonics 2, 4, 5, and 7. The harmonic components of the complex tones were added in sine phase. Inverted triangles show thresholds for a 570-Hz sine-phase complex tone; triangles show thresholds for 570-Hz random-phase complex tones. Modified from Figures 4, 5, and 7 of Lohr and Dooling (1998), with the authors' permission. © 1998 by the American Psychological Association. Adapted with permission.

In a related experiment, Cynx et al. (1990) examined discrimination in zebra finches when a harmonic component was missing. This experiment is actually more like a profile analysis experiment (e.g., Green and Kidd 1983) than a mistuned harmonic experiment. Zebra finch calls were comprised of a F0 of 615 Hz and higher harmonic components (up to the 9th harmonic). In these experiments, zebra finches were trained in a Go/No Go task to discriminate the normal call from a call in which the 2nd harmonic component had been removed. Birds were then tested with normal calls, calls that were missing the 2nd harmonic, and calls in which other harmonics were missing (e.g., missing fundamental, missing 3rd harmonic, etc.). Half of the birds were trained to respond when the stimulus was the normal call and not to respond to the call with the missing 2nd harmonic; these birds showed behavioral responses to all stimuli, except the call with the missing 2nd harmonic. The other half of the birds were trained to respond to the call with the missing 2nd harmonic and not to respond to the normal call; these birds showed behavioral responses only to the call with the missing 2nd harmonic. Thus, the results indicate that the presence or absence of the 2nd harmonic controlled the behavioral response and suggest that zebra finches were able hear the 2nd harmonic as a separate sound source. In a similar experiment, Lohr and Dooling (1998) showed that the thresholds for detecting a decrease in amplitude of a particular harmonic in a tone complex were not significantly different among zebra finches, budgerigars, and human listeners.

As described previously, a mistuned harmonic can be heard as a separate sound source from the harmonic background; this ability to hear out an individual component is a form of analytic listening. Fay (1992a) conducted a study of analytic listening in goldfish using two-tone complexes. Goldfish were trained to respond to a single tone at either 166 Hz or 724 Hz and then were tested in a stimulus generalization paradigm using single tones. Generalization gradients showed a monotonic decrease in behavioral response when the frequency of the test tone varied from the frequency of the training tone. The gradient in behavioral responses is consistent with the hypothesis that goldfish possess a perceptual dimension along the physical dimension of frequency. However, when goldfish were trained to respond to a two-tone complex comprised of frequencies at 166 Hz and 724 Hz and tested with single tones, a bimodal generalization gradient was obtained with maximum responses at 166 Hz and 724 Hz. These findings suggest that goldfish can hear out the individual frequency components of the tone complex.

### *3.5 Rippled Noise Processing*

Rippled noises are generated when wideband noise is delayed ( $T$  ms), attenuated, and the delayed noise is added to (or subtracted from) the original, undelayed version of the noise. Each successive delay and add operation is referred to as an iteration. Thus, if the rippled noise is delayed again and added to the original, undelayed noise, then the output is a rippled noise of two iterations. A

rippled noise having infinite iterations can be achieved by adding the delayed noise to the original noise through a positive feedback loop. Rippled noises of one iteration have been called cosine noises, whereas rippled noises of infinite iterations have been called comb-filtered noises.

Rippled noises are named as such because their spectra are rippled along the frequency axis. When the delayed noise is added to the undelayed noise, the spectrum of rippled noise shows peaks at integer multiples of  $1/T$ ; this is the harmonic condition for rippled noise. When the delayed noise is subtracted from the undelayed noise, the spectrum of rippled noise shows valleys at integer multiples of  $1/T$  with the peaks occurring at odd integer multiples of  $1/(2T)$ ; this is the inharmonic condition for rippled noise. The spectral peaks are broad for rippled noises of one iteration, but become sharper as the number of iterations increase. As the amount of attenuation in the delay-and-add network is increased, there is a decrease in the peak-to-valley ratio in the rippled spectrum. The waveform autocorrelation functions of iterated rippled noises show positive correlations at time lags corresponding to integer multiples of  $T$  when the delayed noise is added. When the delayed noise is subtracted from the undelayed noise, the waveform autocorrelation functions show alternating negative and positive correlations at time lags corresponding to integer multiples of  $T$ . Autocorrelation functions for rippled noise of one iteration show one positive correlation at the time lag corresponding to the delay for the added condition and one negative correlation at the time lag corresponding to the delay for the subtracted condition. As the amount of attenuation in the delay-and-add network is increased, there is a decrease in the heights of the peaks in the autocorrelation functions.

Rippled noises have become an important set of stimuli for studying pitch perception (see Plack and Oxenham, Chapter 2 for perception in human listeners; Winter, Chapter 4 for neurophysiological responses; Griffiths, Chapter 5 for imaging in humans; de Cheveigné, Chapter 6 for auditory models). In animal behavioral studies, rippled noises have been used as maskers to measure the frequency selectivity of auditory filters, but these studies are not concerned with the processing of rippled noises per se, and thus are not discussed here. Questions concerning the auditory processing of rippled noises have been addressed in three animal studies: goldfish (Fay et al. 1983), chinchilla (Shofner and Yost 1995), and budgerigar (Amagai et al. 1999). In these studies, animals were trained either to discriminate a rippled noise from a flat-spectrum wideband noise (i.e., 'coloration' discrimination) or to discriminate between two rippled noises having different delays (i.e., 'pitch' discrimination). In either case, the amount of attenuation in the delay-and-add network is increased until the animal can just discriminate between the two stimuli.

Figure 3.6 summarizes some of the behavioral data obtained from the studies of rippled noise processing in animals. The figure shows a wide range in the thresholds among the animals and humans studied. The budgerigar appears to be the most sensitive among the animals, having thresholds well within the range of human thresholds, and thresholds appear to be independent of delay. The



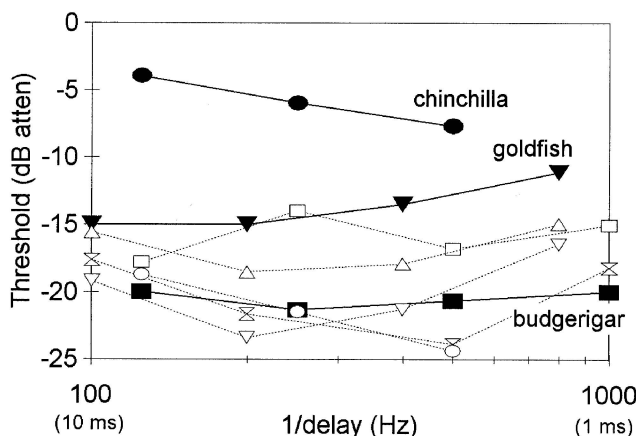


FIGURE 3.6. Thresholds for rippled noise discrimination in animals. Threshold is indicated in terms of the amount of attenuation in the rippled noise delay-and-add network. *Filled triangles* show data from goldfish in a pitch discrimination task (Fay et al., 1983); *filled circles* show data from chinchillas in a coloration discrimination task (Shofner and Yost, 1995); *filled squares* show data from budgerigars in a coloration discrimination task (Amagai et al., 1999). For comparison, *open symbols* show data from human listeners. *Open triangles* show data for coloration discrimination (Bilsen and Ritsma, 1970); *open inverted triangles* show data for pitch discrimination (Bilsen and Ritsma, 1970); *open hourglass* shows data for pitch discrimination (Yost and Hill, 1978); *open circles* show data for coloration discrimination in the same behavioral paradigm used for chinchillas (Shofner and Yost, 1995); *open squares* show data for coloration discrimination in the same behavioral paradigm used for budgerigars (Amagai et al., 1999).

goldfish thresholds appear to be close to the upper limit of human thresholds for longer delays, but deviate from human thresholds as the delay becomes shorter. The thresholds for the chinchilla are clearly above those of human listeners. Although it appears that the chinchilla thresholds decrease as the delay decreases, these authors report that the slope of the threshold versus delay function is not significantly different from 0. Thresholds in chinchillas (Shofner and Yost 1995) and goldfish (Fay et al. 1983) are independent of the overall stimulus level. Shofner and Yost (1995) also report no significant difference in the slopes of the psychometric functions between chinchillas and human listeners. Fay et al. (1983) measured the threshold change in delay when the attenuation was fixed at 0 dB and found a constant Weber fraction of 0.06 with no difference in discrimination for rippled noises produced in the delay-and-add network versus rippled noises produced in the delay-and-subtract network.

One aspect of rippled noise processing for human listeners that has been raised in the literature is that coloration or pitch discrimination could possibly be carried out doing intensity discrimination through a single auditory filter.

That is, subjects could selectively “listen” to just one auditory filter and monitor intensity changes within that filter. To control for this, coloration or pitch discrimination thresholds can be measured when the overall level of the rippled noises is varied randomly among trials. All three of the above animal studies (Fay et al. 1983; Shofner and Yost 1995; Amagai et al. 1999) varied the overall level of the sounds and found no effect on thresholds. Thus, auditory processing of rippled noise stimuli among goldfish, chinchillas, and budgerigars is likely to be accomplished by combining the information about rippled noise in the central auditory system across auditory filters, similar to that described for human listeners.

Shofner and Yost (1995) also compared the performance in chinchillas for ‘coloration’ discrimination of rippled noise for infinite iterations with that of rippled noise for one iteration. It was observed that performance for the discrimination of rippled noise of one iteration with a delayed noise attenuation of 0 dB was similar to the performance for the discrimination of rippled noise of infinite iterations with a delayed noise attenuation of  $-6$  dB. What is interesting about this comparison is that the shapes of the spectra of these two rippled noises are different. The one iterated rippled noise has a spectrum with broad peaks, but large peak-to-valley ratios, whereas for this infinitely iterated rippled noise, the spectral peaks are sharp, but the peak-to-valley ratios are smaller (see Shofner and Yost 1995). However, comparison of the waveform autocorrelation functions shows that the first peak is similar in height for both of these rippled noises. Thus, similar to conclusions about rippled noise processing in human listeners, the results obtained in chinchillas are more consistent with a simple temporal processing mechanism rather than a simple spectral mechanism.

Recently, Shofner (2002) used a stimulus generalization paradigm in order to study the perception of rippled noise stimuli in chinchillas. Chinchillas were trained to discriminate a cosine-phase harmonic tone complex from a wideband noise, and then tested in the generalization paradigm with various iterated rippled noises substituted for the harmonic tone complex. Figure 3.7 shows that the behavioral responses are relatively small to the infinitely iterated rippled noise having a delayed noise attenuation of  $-1$  dB. Of the rippled noises tested, this particular iterated rippled noise generates the most salient pitch in human listeners (see Shofner and Selas 2002). These particular animals had no previous experience listening to iterated rippled noise stimuli. For comparison, Figure 3.7 also shows the psychometric functions obtained from the discrimination task (Shofner and Yost 1995); these animals were trained to discriminate iterated rippled noise from wideband noise and received positive reinforcement for correct behavioral responses to iterated rippled noise stimuli having delayed noise attenuations ranging from  $-1$  dB to  $-8$  dB. Clearly, the chinchillas in the discrimination experiment are attending to different cues than the animals in the generalization experiment. Note that one animal (C7), which participated in both the discrimination and generalization experiments, showed a difference in the generalization gradients with most other animals. These results suggest that there may be a difference in listening strategy between animals with and without previous experience listening to stimuli like iterated rippled noises.

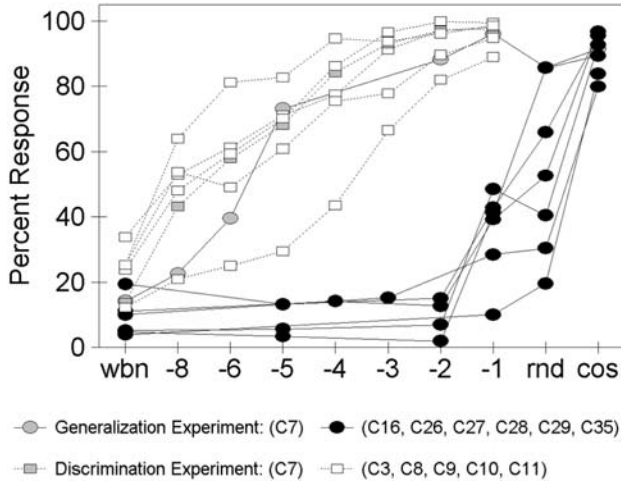


FIGURE 3.7. Comparison of behavioral performance between two groups of chinchillas in either a discrimination task or a generalization task. *Filled circles* show data from the stimulus generalization experiment; *open squares* show data from the discrimination experiment. *Gray circles and squares* indicate data from the only animal (C7) to participate in both experiments. Stimuli are labeled on the *x*-axis as follows: wideband noise (wbn); infinitely iterated rippled noise having delayed noise attenuations of  $-8$ ,  $-6$ ,  $-5$ ,  $-4$ ,  $-3$ ,  $-2$ ,  $-1$  dB; random-phase harmonic tone complex (rnd); cosine-phase harmonic tone complex (cos). Complex tones and iterated rippled noises have periods and delays of 4 ms. Note that although the *x*-axis shows a nominal scale, there is a general increase in stimulus periodicity strength for stimuli from *left to right* along the *x*-axis (see Shofner 2002). The lines connect the data points for each individual animal. In the discrimination experiment, chinchillas were trained to discriminate infinitely iterated rippled noise with  $-1$  dB attenuation from wideband noise and tested with the other rippled noises. In the generalization experiment, chinchillas were trained to discriminate cosine-phase harmonic tone complex from wideband noise and tested with the other stimuli shown. Discrimination data are from Shofner and Yost (1995); generalization data are from Shofner (2002).

Finally, two additional discrimination studies using rippled noises should be noted. Rippled noise processing has also been studied in the dolphin (*Tursiops truncatus*) (Au and Pawloski 1989). The delays used in this study ranged from 0.01 to 1 ms, and the dolphin showed its best performance for delays around 0.1 ms. Most of this delay range falls far outside of the range of delays associated with rippled noise pitch in human listeners (i.e., 1 to 10 ms). Amagai et al. (1999) also studied rippled noise processing in the budgerigar using sinusoidally spaced ripples on a logarithmic scale. Although log-rippled noises may not be as relevant for pitch perception as the linear-rippled noises discussed above, they have become important for approaches based on a linear-systems analysis to auditory processing. ‘Coloration’ discrimination thresholds for budgerigars using log-rippled noise were lower than those of human listeners tested in the same procedure.

### 3.6 *Dominance Region*

Psychophysical studies in human listeners have shown that not every frequency region of a complex, harmonic sound is equally important in generating the perceived pitch of the sound. It is typically cited that the frequencies in the region of the 3rd to 5th harmonics are most effective in evoking the pitch. This effect of frequency location or harmonic number on pitch is known as spectral dominance or the dominance region (see Plack and Oxenham, Chapter 2).

Two studies concerning the spectral dominance region have been carried out in animals, and both of these studies have employed bandpass-filtered rippled noise as stimuli. Au and Pawloski (1989) showed no evidence for a dominance region in the dolphin for bandpass-filtered rippled noises at a delay of 0.1 ms. As stated above, this delay is far outside of the delays that give rise to rippled noise pitch in human listeners.

Shofner and Yost (1997) did find evidence for a dominance region for ‘coloration’ discrimination by the chinchilla. Figure 3.8A shows behavioral performance at center frequencies corresponding to integer multiples of the peaks in the iterated rippled noise. The functions illustrated clearly have a bandpass characteristic, and it can be observed in this example that for any given delayed noise attenuation, behavioral performance was best around center frequencies of 750 to 1000 Hz. These center frequencies correspond to 3rd and 4th harmonics of a 4-ms delay rippled noise. Similar results have been described in human listeners for rippled noise stimuli (Yost and Hill 1978; Yost 1982; Leek and Summers 2001). Figure 3.8B compares the ‘coloration’ discrimination thresholds for chinchillas (Shofner and Yost 1997) with human thresholds (Leek and Summers 2001) obtained under similar conditions. It should be noted that the location of the dominance region in chinchillas varies with the corresponding pitch (i.e.,  $1/T$ ) of the rippled noise. That is, there is a trend in the data showing that the dominance region is centered around the 5th to 6th harmonics for rippled noises of long delays (i.e., 8 ms), but is centered around the 2nd to 3rd for rippled noises of shorter delays (i.e., 2 ms). Thus, for rippled noises at short delays, the lower harmonics are dominant, whereas for rippled noises of longer delays, the higher harmonics are dominant. Similar trends have been described for the dominance regions of rippled noises and complex tones in human listeners (see Plack and Oxenham, Chapter 2).

### 3.7 *Phase Effects on Periodicity Discrimination and Perception*

The effects of starting phase of the individual frequency components in a complex tone have been studied extensively for pitch perception in human listeners (see Plack and Oxenham, Chapter 2). In several of the animal studies discussed above, the effects of starting phase were also examined, and those results are presented in this section.

Although phase effects were not studied as part of the experiments describing

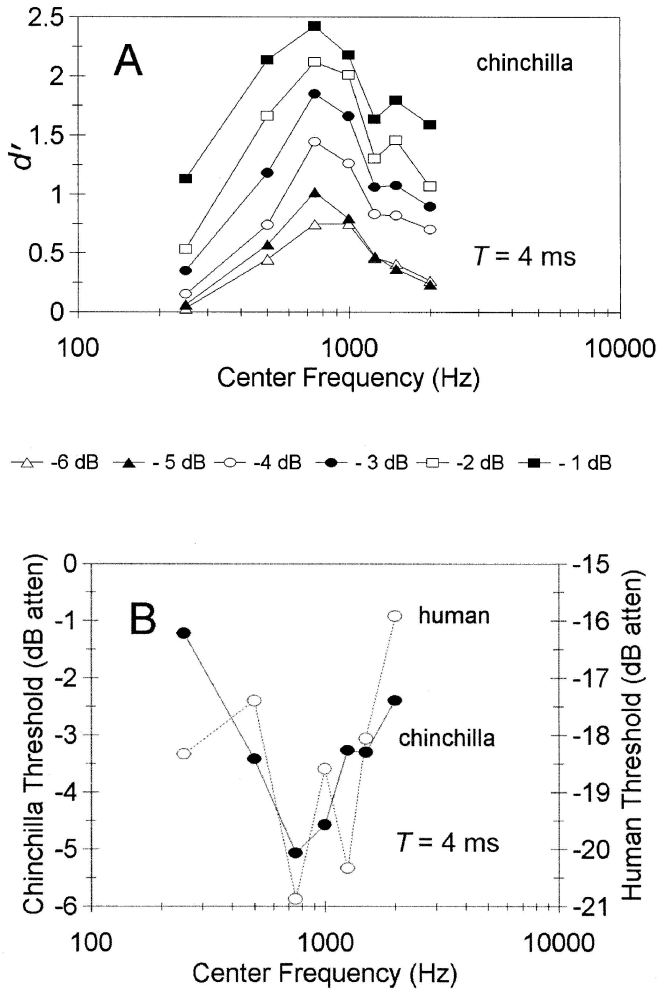


FIGURE 3.8. (A) Behavioral performance as a function of center frequency for chinchillas for bandpass filtered rippled noises of a fixed delay of 4 ms. Performance is measured as  $d'$  in a coloration discrimination task. Averaged data are from Shofner and Yost (1997). Symbols indicate the delayed noise attenuation in dB. Moving vertically along the y-axis at a fixed center frequency moves along the psychometric function for that center frequency. (B) Discrimination threshold as a function of center frequency for chinchillas (filled circles) and human listeners (open circles). Chinchilla data are from Shofner and Yost (1997); human data are from Figure 4 of Leek and Summers (2001) with the authors' permission. Chinchilla thresholds were defined as the delayed noise attenuation that would result in a  $d' = 1$ . The delay of the bandpass filtered rippled noise is 4 ms. The human function has been displaced by 15 dB to facilitate the comparison. The bandpass filters used in both the chinchilla and human studies were one octave wide.

the missing fundamental in animals, phase effects have been studied in frequency discrimination of complex tones. Shofner (2000) trained chinchillas to discriminate the F0s of complex tones comprised of the F0s and the 2nd through 10th harmonics with individual components added in either cosine-starting phase or in random-starting phase. Thus, the stimuli were comprised primarily of the resolved, low-frequency harmonics. The cosine- and random-phase tone complexes have identical waveform autocorrelation functions, but different envelope autocorrelation functions. Animals were trained to discriminate the tone complex with a 250-Hz F0 from a tone complex having a higher F0. The psychometric functions for the random-phase tone complexes were similar to those obtained for the cosine-phase tone complexes, and there was no significant difference in the mean discrimination thresholds between cosine- and random-phase tone complexes. This finding is similar to results observed from human listeners with normal hearing for complex tones comprised of the first 12 harmonics (Moore and Glasberg 1988; Moore and Peters 1992).

Lohr and Dooling (1998) examined the effect of starting phase on the detection of mistuned harmonics in zebra finches. Stimuli were harmonic tone complexes comprised of the first 16 harmonics of a 570-Hz F0 in which all components were added in sine or random starting phases. Mistuning detection thresholds were significantly higher for human listeners than for zebra finches. There were no significant differences between thresholds for sine-phase and random-phase complexes for human listeners (see open triangles in Fig. 3.5). For zebra finches, the thresholds for random-phase complexes were significantly higher than those for the sine-phase condition (see gray triangles in Fig. 3.5), suggesting that birds may be more sensitive to phase than human listeners.

Several studies have examined phase discrimination per se in mammals, birds, and anuran amphibians. In a study similar to that in human listeners by Mathes and Miller (1947), monkeys were trained to discriminate quasi-frequency modulated tones (QFM tones) from SAM tones (Moody et al. 1998). In SAM tones, the phase of the center frequency is 0°, whereas in QFM tones, the phase of the center frequency is 90°. The envelope of the stimulus varies from being relatively flat (QFM) to highly modulated (SAM). Psychometric functions were generated as the starting phase of the center frequency was systematically varied from 90° to 0°. For a fixed center frequency, phase discrimination thresholds generally decreased (i.e., smaller phase changes were detectable) as modulation frequency increased. For a fixed modulation frequency, phase discrimination thresholds showed no systematic change with center frequency.

Bullfrogs, but not green treefrogs, appear to be sensitive to starting phase. Large evoked vocal responses were obtained from bullfrogs to synthetic mating calls in which harmonic components were added in cosine phase or random phase, but significantly smaller evoked calling responses were obtained for synthetic mating calls in which components were added in alternating starting phase (Hainfeld et al. 1996; Simmons et al. 2001). It should be emphasized that the F0 of these synthetic calls was fixed at the F0 of the natural call. Evoked calling

in the green treefrog, however, was not affected by these same phase manipulations (Simmons et al. 1993).

Dooling et al. (2002) studied phase discrimination in budgerigars and human listeners. Stimuli were harmonic tone complexes comprised of the F0 and all components up to and including 5000 Hz. Harmonic components were added either in cosine starting phase or random starting phase. Figure 3.9 summarizes the average discrimination data between human listeners and budgerigars. Discrimination performance was similar between budgerigars and human listeners at a F0 of 200 Hz, but budgerigars showed significantly better performance than human listeners as the F0 increased. Both functions show a lowpass characteristic, but the cutoff frequency for the budgerigars is much higher than for human listeners.

Dooling et al. (2002) also studied the discrimination between positive- and negative-Schroeder-phase harmonic tone complexes in three species of birds and human listeners. Positive-Schroeder-phase tone complexes are generated by having a monotonic increase in the phase of the harmonic components, whereas the starting phase decreases monotonically for negative-Schroeder phase tone complexes. The stimuli have identical power spectra and differ only in their phase spectra. Figure 3.9 summarizes the average discrimination data between

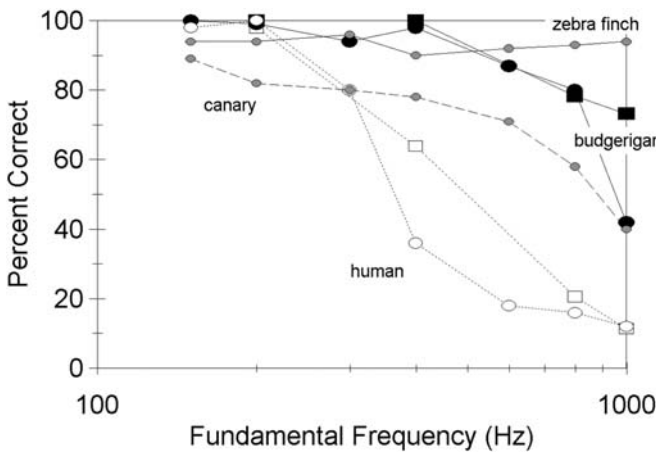


FIGURE 3.9. Phase discrimination performance for complex tones as a function of F0 in birds and humans. Behavioral performance is measured as percent correct. *Black symbols* show data from budgerigars; *gray symbols* show data from zebra finches and canaries; *open symbols* show data from human listeners. *Squares* show data for discrimination of cosine- from random-phase harmonic tone complexes; *circles* show data from discrimination of positive- from negative-Schroeder-phase harmonic tone complexes. Modified from Figures 2 and 5D of Dooling et al. (2002), with the authors' permission.

human listeners and three species of birds. There was a similarity in behavioral performance between human listeners and birds at low F0s, but birds showed significantly better discrimination of positive- from negative-Schroeder-phase tone complexes as the F0 increased. Zebra finches showed higher performance at a 1000-Hz F0 than either budgerigars or canaries (*Serinus canaria*) (Fig. 3.9). These data indicate that birds are highly sensitive to phase.

In human listeners, starting phase can have an effect on pitch strength. A random-phase harmonic tone complex can have a slightly weaker pitch strength than a cosine-phase harmonic tone complex (e.g., Lundeen and Small 1984; Shofner and Selas 2002). In the generalization experiment previously described in which chinchillas were trained to discriminate cosine-phase harmonic tone complexes from wideband noise, Shofner (2002) also tested chinchillas using random-phase tone complexes. Chinchillas typically gave smaller behavioral responses to random-phase harmonic complex tones than to cosine-phase tone complexes (compare rnd versus cos in Fig. 3.7). The average behavioral response (in terms of percent generalization) to the random-phase tone complexes was 49% compared to 90% for the cosine-phase tone complex. This decrease in behavioral response with starting phase in chinchillas is in contrast to the results obtained using a scaling procedure in human listeners with the identical stimuli. Human listeners judge the pitch strengths of these random-phase and cosine-sine phase harmonic tone complexes to be 93% and 99%, respectively (Shofner and Selas 2002). The results suggest that the temporal information in the stimulus envelope has a large effect on the perception in chinchillas (Shofner 2002), whereas the temporal information in the fine structure has a large effect on the perception in human listeners.

#### 4. Music Discrimination and ‘Pitch’ Perception

Pitch, timbre, and rhythm are important percepts in the appreciation of music by human listeners. Pitch is essential for melody recognition in human listeners, and indeed one yardstick that has been used by psychoacousticians to determine if a synthetic sound evokes a pitch percept is whether or not a melody can be recognized with the sound. Many of the issues that relate to pitch and music perception have also been studied in animals. Both pigeons (Porter and Neuringer 1984) and carp (*Cyprinus carpio*, Chase 2001) can discriminate among complex musical sequences, and both appear to show categorical perception for musical stimuli. These experiments used sampled sequences from previously recorded instrumental music as stimuli, and therefore, perceptions of ‘pitch,’ ‘timbre,’ and ‘rhythm’ would all have been available as potential discrimination cues for the animals. Poli and Previde (1991) trained rats to discriminate the melody “Frere Jacques” from a rearranged version of the melody in which both note duration and melody rhythm were maintained. Melodies were recorded using either trumpet or guitar. Rats were able to discriminate among the mel-



odies, but it was concluded that ‘timbre,’ not ‘pitch,’ was the perceptual cue used for the discrimination.

### 4.1 Discrimination of Chords

In the literature of music perception, a tone is defined as a complex sound comprised of a F0 and its related harmonics. That is, a musical tone is really a harmonic tone complex. Thus, the musical tone C<sub>4</sub> (i.e., middle C) is a tone complex comprised of a 262-Hz F0 and some number of higher harmonics. A chord is then comprised of two or more of these complex tones. For example, the C major chord (C<sub>4</sub>–E<sub>4</sub>–G<sub>4</sub>) is comprised of three separate harmonic complex tones having F0s of 262 Hz (C<sub>4</sub>), 330 Hz (E<sub>4</sub>), and 392 Hz (G<sub>4</sub>).

Hulse et al. (1995) trained starlings to discriminate between two different chords in which each chord was comprised of three musical tones. The chords differed in the intervals between musical tones: the intervals of one chord followed structures commonly used in music (i.e., C–E–G), whereas the intervals of the other chord did not (i.e., C–D–G). Starlings learned to discriminate these chord differences and transferred the discrimination (i.e., generalized) to new F0s of the root musical tones. This finding suggests that starlings show perceptual invariance for chord structure. That is, chords of similar interval structure are perceived as being similar regardless of F0 of the root musical tone. Hulse et al. (1995) also found that starlings trained to discriminate two chords showed similar performance when tested with inverted chords. The first inversion of a chord occurs when the root tone is increased in frequency by one octave. For example, the first inversion for the C major chord C<sub>4</sub>–E<sub>4</sub>–G<sub>4</sub> becomes E<sub>4</sub>–G<sub>4</sub>–C<sub>5</sub>. Thus, starlings also showed perceptual invariance for chord inversions. Hulse et al. (1995) argued that although starlings could base chord discrimination on the interval structure of the chords, the discrimination could also be based on the relative consonance of the chords.

Izumi (2000) trained Japanese monkeys (*Macaca fuscata*) in a Go/No Go task to discriminate between two-note chords on the basis of consonance. Monkeys were presented simultaneously with two musical tones (with each musical tone being comprised of six harmonics) and were trained to respond when the frequency ratio (consonance) changed from an octave (1:2) to a major seventh (8:15). That is, monkeys were trained to respond when the two-note chords changed from being consonant to being dissonant. If monkeys based their discrimination on the frequency ratio or musical interval (i.e., chord structure), then behavioral performance for discriminating new *consonant* chords should be poor, whereas performance for discriminating new *dissonant* chords should be high. However, if monkeys based their discrimination on the absolute frequencies of the chords, then the frequency ratio or musical interval should not have an effect on discrimination, regardless of whether the new chords were consonant or dissonant. When tested with 14 new chords of varying frequency ratios, monkeys showed good discrimination for dissonant chords and poor discrimi-

nation for consonant chords (i.e., octave and unison chords), suggesting that the discrimination was based on chord structure (i.e., consonance). The results of these two studies suggest that the perception of consonance is not unique to human listeners.

## *4.2 Perception of Frequency Contours and Octave Generalization*

Many of the above studies that have been described in this overview use stimuli that are complex in the spectral domain. In contrast, the following discussion describes studies that have used stimuli that are complex in the time domain (i.e., complex patterns of single tones). One aspect concerning melody recognition in human listeners that is related to pitch perception is the frequency contour of the melody. The frequency contour refers to the changes in the sequential pattern of the pitches of the notes; that is, to changes in the successive intervals between pitches of individual notes. For example, in human listeners, changing the frequency of individual notes in octave steps does not affect the recognition of the melody, and the preservation of a recognizable melody with octave transpositions in the frequencies of the individual notes is known as octave generalization. The discrimination of frequency contours and octave generalization have been studied in animals, and important to this discussion are the concepts of absolute and relative 'pitch' perception in animals. Absolute 'pitch' refers to basing the discrimination on the absolute frequencies of the tones, whereas relative 'pitch' refers to basing the discrimination on the intervals between individual frequencies of the tones. Thus, to demonstrate relative 'pitch' perception, the animal must continue to respond to test sequences in which the tonal frequencies have been transposed up or down by some fixed interval.

Octave generalization, then, would be an example of relative 'pitch' perception in which the tonal frequencies have been transposed in octave steps. One of the first studies to address octave generalization in animals was carried out by Blackwell and Schlosberg (1943). Rats were trained to respond during the presentation of a 10-kHz single tone and tested in a stimulus generalization paradigm with single tones of varying frequencies. Large behavioral responses were obtained for the 10-kHz tone, and the behavioral response decreased systematically as the frequency of the test tone decreased from 10 kHz (i.e., there was a gradient in behavioral responses). However, there was an increase in behavioral responses to 5-kHz tones, which are one octave below the 10-kHz training tone. Because of this peak in the generalization gradient that occurred at 5 kHz, these authors concluded that rats show octave generalization. However, this finding may have been due to harmonic distortion, and the results have been difficult to replicate using modern technology. Generalization gradients for starlings (Cynx 1993) and for goldfish (Fay 1970) also show monotonic decreases in behavioral response as the test tone frequencies vary from the training frequency, but with no deviation from the gradient at octave frequencies.

Hulse et al. (1984) trained starlings to discriminate between two four-tone sequences: in one sequence frequencies increased and in the other sequence frequencies decreased. The frequencies of the sequences were within a one-octave range. Starlings maintained the discrimination when the intensity of the tones varied, indicating that intensity was not a cue. Hulse et al. (1984) concluded that the discrimination depended partly on absolute 'pitch' cues and partly on relative 'pitch' cues. In later studies (Hulse and Cynx 1985; Cynx et al. 1986), starlings were trained to discriminate the ascending and descending 'pitch' sequences as in the previous study, but were then tested in a generalization task in which the absolute frequencies of the test sequences were either lowered or raised by one octave. Starlings immediately lost the discrimination when the tones were shifted to either lower or higher octaves; that is, they failed to show octave generalization. However, starlings continued to generalize when tested with novel tone sequences in which the frequencies were shifted within the frequency range of the training exemplars. Similar results have been obtained with cowbirds (*Molothrus*) and mockingbirds (*Mimus*, Hulse and Cynx 1985) and zebra finches and pigeons (Cynx 1995). It is interesting to note that pigeons, a nonsongbird species, required more trials to learn the initial discrimination than did songbirds (Cynx 1995). Other studies in the starling (Hulse and Cynx 1986; Page et al. 1989; MacDougall-Shackleton and Hulse 1996) have addressed questions regarding the salience of absolute and relative 'pitch' perception and have concluded that whereas starlings possess both absolute and relative 'pitch' perception, it is absolute 'pitch' that appears to be most salient in starlings. That is, starlings seem to have a predisposition for absolute 'pitch' perception. Budgerigars also are not sensitive to the frequency contour of a sequence of tones, but respond to the absolute frequencies in the sequences (Dooling et al. 1987a), and when the frequency content of natural vocal calls of budgerigars is altered, classification of the calls based on a multidimensional scaling analysis showed that absolute 'pitch' is a salient perceptual dimension (Dooling et al. 1987b).

D'Amato and colleagues have studied the perception of frequency contours in cebus monkeys (*Cebus apella*) and rats. In these experiments, frequency contours were generated as sequences of single tones and animals were tested in a Go/No Go procedure. D'Amato and Salmon (1982) demonstrated that monkeys can discriminate a tune (sequence of single tones that both ascended and descended in frequency forming a random melody) from a glissando (sequence of tones that either ascended or descended in frequency). Transposing the frequencies of the tune by one octave had essentially no effect on behavioral performance, whereas a two-octave transposition impaired performance. Similar results were found in rats, except rats showed higher performance for two-octave transpositions than did monkeys, and rats learned the behavioral task faster than monkeys. These results suggested that monkeys and rats can discriminate tone patterns based on the frequency contours (i.e., on the structures of the tunes), but subsequent studies by D'Amato and co-workers (see below) failed to support this conclusion.

These findings were extended by D'Amato and Salmon (1984), again using tunes generated by single tones. Tune 1 was comprised of 10 monotonically decreasing tones having a mean frequency of 2902 Hz; tune 2 was a highly structured sequence of alternating low- to high-frequency tones in which the overall frequency increased and the mean frequency was 898 Hz. Monkeys and rats showed a high level of discrimination performance between these two different tunes, but because of the difference in mean frequencies, discrimination could be based on the difference between the mean frequencies rather than the frequency contours. Again, rats learned this discrimination faster than monkeys. Monkeys and rats also showed a high level of discrimination performance when they were tested using randomized versions of the same two tunes. In the randomized condition, the frequency contours of the tunes were different from the previous condition, but the mean frequencies were still 2902 Hz and 898 Hz. These findings argue that the discrimination was based on the overall frequency difference (i.e., mean absolute frequencies) rather than based on the structure of the tone sequences. D'Amato and Salmon (1984) also studied the discrimination of two tunes, each having a distinct pattern of tones, but having similar mean frequencies. Both monkeys and rats were able to discriminate these two tunes with a high level of behavioral performance. Behavioral responses for both monkeys and rats decreased when the frequencies of the tones making up the tunes were lowered by one octave, suggesting that octave generalization did not occur. D'Amato and Colombo (1988) also found no evidence that monkeys could discriminate tone patterns based on the frequency contours (i.e., relative 'pitch' cues) and concluded that the discrimination was based on the absolute frequencies of the first few tones of the sequences.

Although the tone patterns used in the above studies by D'Amato and colleagues were structured, the frequency intervals between tones were generally not fixed. Izumi (2001) studied the perception of frequency contours in the monkey in which the intervals of the tones in the sequences were fixed at two semitones (i.e., 1/6 octave interval). Monkeys were trained in a Go/No Go paradigm to discriminate falling three-tone sequences from rising three-tone sequences. During training, four possible sets of rising and falling sequences were used that covered a range of frequencies from 440 Hz to 1108 Hz. Monkeys were then tested in a generalization task with three different sets of probe sequences comprised of three rising and falling tones. For the first probe sequence, the frequencies of the tones fell outside of the range of tone frequencies for the training sequences. In this case, the behavioral responses to the probe sequences were higher than those for the training sequences, suggesting that monkeys based the discrimination on the absolute frequency differences between the probe and training stimuli. Monkeys were also tested using three-tone sequences in which the specific sequence of the three tones was not one of the training sequences, but in which the frequencies of the tones in the probe sequence fell within the frequency range of the training sequences. In this case, the behavioral responses for the training and probe sequences were similar, sug-

gesting that monkeys based the discrimination on the frequency contours (i.e., relative differences between the individual tones in the sequence). Similar behavioral results were obtained when the intervals in the probe sequence were larger than those in the training sequences. Izumi (2001) concluded that when the probe frequencies are within the range of training frequencies, then monkeys base the discrimination on relative 'pitch' differences, but when the probe frequencies are outside of the range of training frequencies, monkeys base the discrimination on absolute 'pitch' cues. This conclusion is similar to that previously described for birds.

The above discussion indicates that animals can discriminate tone sequences based on salient absolute 'pitch' cues, and that relative 'pitch' cues are less salient in animals. In general, the preceding studies have used tone sequences in which the frequencies were essentially random. Wright et al. (2000) have argued that "contour and octave generalization should depend on relating two musical passages," and relating the melodies of two musical passages or tunes is a same-different concept that cannot be easily applied to the typical Go/No Go paradigms used in animal psychophysical experiments. These authors first used a variety of natural and environmental sounds to train Rhesus monkeys on the same-different concept. Monkeys were then trained and tested in a series of generalization experiments using melodies as stimuli. The experimental conditions and behavioral results are summarized in Table 3.1 and Figure 3.10, respectively. In Figure 3.10, the open bars indicate the behavioral responses to training melodies; note that these indicate high levels of response. The filled bars indicate behavioral responses to test melodies. High levels of response to the test melodies indicate that the animals have transferred the discrimination or generalized to the new stimulus, but low levels of response indicate that the animal has not generalized to the new stimulus.

In experiment 1, monkeys were trained in the same-different task to respond to six-note random-synthetic melodies and then tested in the generalization task using the same melodies transposed in frequency within a four-octave range. Figure 3.10 indicates that monkeys showed little generalization to the transposed melodies; this finding is similar to the results of D'Amato and colleagues previously described. However, when monkeys were trained to respond to melodies comprised of six-notes of childhood songs (experiment 2), generalization to test melodies occurred when the same melodies were transposed up in frequency by one octave. That is, monkeys demonstrated octave generalization to childhood song melodies. When monkeys were again retested using the random-synthetic melodies (experiment 3), no generalization occurred to transposed melodies. This result indicates that the octave generalization that occurred for the childhood songs in experiment 2 was not related to experience, but rather to some difference between the childhood songs and random melodies. However, since the frequency transpositions for the random melody experiments were not octave transpositions, then experiment 4 used six-note random-synthetic melodies that were transposed up in frequency by one octave. Again, monkeys failed to gen-

TABLE 3.1. Summary of experimental conditions for data in figure 10 from Wright et al. (2000).

| Experiment | Training condition                                                      | Testing condition                                            |
|------------|-------------------------------------------------------------------------|--------------------------------------------------------------|
| 1          | Six-note random-synthetic melodies                                      | Same melodies transposed within a four-octave range          |
| 2          | Six notes of childhood melodies (12 different songs)                    | Same melodies transposed up one octave                       |
| 3          | Replication of experiment 3                                             | Replication of experiment 3                                  |
| 4          | Six-note random-synthetic melodies                                      | Same melodies transposed up one octave                       |
| 5.1, 5.2   | Six notes of childhood melodies                                         | Same melodies transposed up one or two octaves, respectively |
| 6.1, 6.2   | Individual notes from childhood songs                                   | Same notes transposed up one or two octaves, respectively    |
| 7a, 7t     | Seven-note atonal or tonal melodies generated with a tonality algorithm | Same melodies transposed up one octave                       |

eralize to the octave-transposed stimuli (Fig. 3.10). Monkeys demonstrated octave generalization for both one- and two-octave transpositions of childhood song melodies (experiments 5.1 and 5.2 in Fig. 3.10), but failed to generalize to the new melodies when they were transposed up in frequency by  $\frac{1}{2}$  or  $1\frac{1}{2}$  octaves. The failure to generalize to  $\frac{1}{2}$  and  $1\frac{1}{2}$  octave transpositions, while showing generalization to whole-octave transpositions, indicates that chroma (i.e., key) is important for octave generalization in monkeys. In experiment 6, monkeys were trained using only the individual notes from the melodies, rather than the melodies themselves. Monkeys failed to generalize when the individual notes were transposed up in frequency by one (experiment 6.1 in Fig. 3.10) or two octaves (experiment 6.2 in Fig. 3.10). Finally, experiment 7 examined whether musical tonality is an important property for octave generalization. In this experiment, seven-note atonal or seven-note tonal melodies were generated using an algorithm that could maximize the tonality of the melodies. In experiment 7, monkeys failed to show octave generalization when atonal synthetic melodies were used (experiment 7a in Fig. 3.10), but did show octave generalization when tonal synthetic melodies were used (experiment 7t in Fig. 3.10).

The results of the series of experiments by Wright et al. (2000) indicate that octave generalization occurred only for childhood songs and for melodies possessing strong musical tonality, and not for individual notes or random melodies, which are characterized by weak musical tonality. That is, the strong tonality makes childhood songs more musical than random melodies, resulting in the formation of a type of gestalt for songs that is not formed for random melodies. Moreover, Wright et al. (2000) argue that since the formation of the gestalt for melodies with strong musical tonality occurs in monkeys, the perception of tonality is not uniquely a human perception.

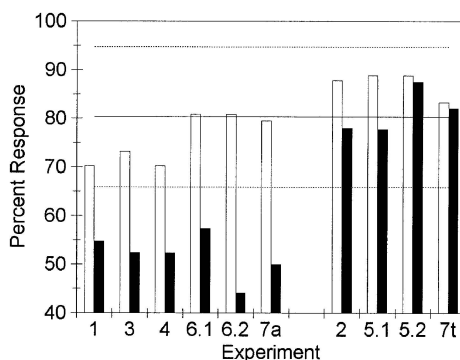


FIGURE 3.10. Bar graph summarizing some of the behavioral data on octave generalization in monkeys from Wright et al. (2000). Percent response indicates the percent of “same” responses. *Open bars* indicate behavioral responses to training stimuli; *filled bars* indicate responses to test stimuli that have been transposed in frequency. *Horizontal solid* and *dotted lines* indicate the average responses  $\pm 2$  standard deviations for training stimuli across all experiments indicated. Chance performance is at 50%. Note that the *filled bars* on the left-hand side fall well under the average behavioral response to the training stimuli and are close to 50%, whereas the *filled bars* on the right-hand side fall close to the average response to the training stimuli. Modified from Figures 2–8 of Wright et al. (2000), with the authors’ permission. © 2000 by the American Psychological Association. Adapted with permission. Table 3.1 indicates the conditions of the experiments.

## 5. Auditory Streaming and ‘Pitch’ Perception

When human listeners are presented with a pattern of sounds (i.e., ABABAB), the individual sounds may be grouped together to form two separate auditory streams (i.e., A–A–A– or B–B–B). When the sounds are comprised of tone sequences, pitch is a perceptual cue that plays an important role in grouping sounds into auditory streams (see Darwin, Chapter 8). As described earlier, animals appear to possess a perceptual dimension related to tone frequency, and animals can perceive differences between tone sequences.

Izumi (2002) has shown that Japanese monkeys trained to discriminate tone sequences of rising frequency from sequences of nonrising frequency of tones (“target” condition) showed no degradation in performance when this same target condition was simultaneously presented with rising and nonrising distractor sequences in which the frequencies used in the distractor sequences were outside of the frequency range used for the target sequences. However, when the frequency ranges for the target and distractor sequences did overlap, the monkeys could no longer discriminate the target sequences. In a similar experiment, MacDougall-Shackelton et al. (1998) trained starlings to discriminate between the tone pattern AAA\_AAA\_AAA\_ . . . from two isochronous patterns using the

same single-tone frequency. These patterns were  $\_A\_A\_A\_ \dots$  and  $A\_A\_A\_A\_A\_ \dots$ . The behavioral responses of starlings were then measured to probe sequences having the pattern  $ABA\_ABA\_ABA\_ \dots$ , where tones A and B differ in frequency. When the differences in frequency of A and B were small (i.e., 50 Hz), the behavioral responses to the  $ABA\_$ sequence were similar to those of the  $AAA\_$ sequence, suggesting that the starling did not segregate the two tones into individual streams. However, when the frequency differences between A and B were large (i.e., 3538 Hz), the behavioral responses to the  $ABA\_$ sequence were similar to those of the isochronous sequence, suggesting that the starling was able to segregate the two tones into individual streams. The results of the above studies are consistent with the hypothesis that animals can also segregate tone patterns into auditory streams. Similar results have also been found in starlings using harmonic complex tones (Braaten and Hulse 1993) and in goldfish using Gaussian-filtered tone pulses (Fay 1998, 2000).

## 6. Summary and Concluding Remarks

This chapter described the behavioral data in vertebrate animals that are related to the perceptual attributes of human pitch perception. In many aspects, ‘pitch’ perception in animals is qualitatively similar in to human pitch perception. For example, behavioral evidence suggests that animals possess a perceptual dimension related to tone frequency; possess a perception of the missing fundamental; possess a spectral dominance region; are able to discriminate rippled noises and detect mistuned harmonics; and show octave generalization and auditory streaming. Thus, the general perceptual processes and neural mechanisms underlying pitch perception do not appear to be unique or special mechanisms that are specific to human listeners. In other words, it can be concluded that the human perceptual processes reflect general vertebrate mechanisms. Given that the basic neural pathways of the central auditory systems are conserved across vertebrates, it is not surprising that similarities can be found in the behavioral data across vertebrates. There are differences, however, in the behavioral data among animals and human listeners, and these differences are perhaps more interesting than the similarities in understanding the biological basis of pitch perception. The following sections are meant to bring forward a few issues that may be important to consider when attempting to relate animal ‘pitch’ perception to human pitch perception.

### *6.1 Issue 1: Peripheral Representations of Complex, Periodic Sounds*

One issue in human pitch perception concerns whether resolved and unresolved frequency components are processed by separate mechanisms or by a single mechanism (Plack and Oxenham, Chapter 2). Arguments for or against separate



mechanisms are still being debated, more than 150 years since the original experiments of Seebeck and Ohm (see Houtsma 1995 for historical review). How might this issue of resolved and unresolved components relate to 'pitch' perception in animals?

One aspect of this issue may relate in part to the differences in the auditory organs among vertebrates. For example, it is known that in most nonhuman mammals, the cochlea is shorter than in humans (see Echteler et al. 1994 for review). Recently, the issue of differences in cochlear length among nonhuman mammals and humans has been explored in regard to the neural representation of speech sounds in the mammalian auditory nerve (Kieffe et al. 2002; Recio et al. 2002). Based on the cochlear frequency-position function derived by Greenwood (1961, 1990), these authors argue that the frequency difference between formants of a vowel will translate into a smaller position difference along the basilar membrane of nonhuman mammals than along the basilar membrane of humans. What effect could a shorter cochlea have on the representation of complex, periodic sounds?

Consider the positions along the basilar membrane of humans and chinchillas for the harmonic components of a complex tone having an F0 of 250 Hz (Fig. 3.11A). For this complex tone, the 250-Hz and 500-Hz components will be separated by 3.4 mm in the human cochlea (e.g.,  $\Delta X_{500-250 \text{ Hz}}$  in Fig. 3.11A) and 1.9 mm in the chinchilla cochlea based on the function derived by Greenwood. Figure 3.11B illustrates the difference in distance along the basilar membrane between adjacent harmonic components for humans and several common laboratory mammals (i.e., the  $\Delta X$  values between each successive pair of components). Note that the distance between any two neighboring harmonic components is smaller for nonhuman mammals than for humans, and this smaller distance has implications in regard to the number of components that may be resolved or unresolved. If frequency resolution along the basilar membrane is *better* for nonhuman mammals than for humans, then the smaller distance between adjacent components might be offset, such that the number of resolved and unresolved components would be equal among nonhuman mammals and humans. Is there any evidence that frequency resolution along the cochleae of nonhuman mammals is better than in humans?

Auditory filter bandwidths in chinchillas derived from notched-noise and rippled-noise masking are similar to those in humans (Niemiec et al. 1992), and the bandwidths of psychophysical tuning curves are similar among nonhuman mammals and humans (see Fay 1992b for review). These data argue that frequency resolution is *similar* along the cochleae of nonhuman mammals and humans. More recently, however, Shera et al. (2002) have reported data suggesting that human auditory filters are sharper than measured previously and are sharper than those measured in other nonhuman mammals. In addition, the single-tone frequency discrimination data (Fig. 3.1) suggest better frequency resolution in humans than in nonhuman mammals. These data argue that frequency resolution is *poorer* in nonhuman mammals than in humans. Thus, the empirical data certainly do not indicate better frequency resolution in nonhuman

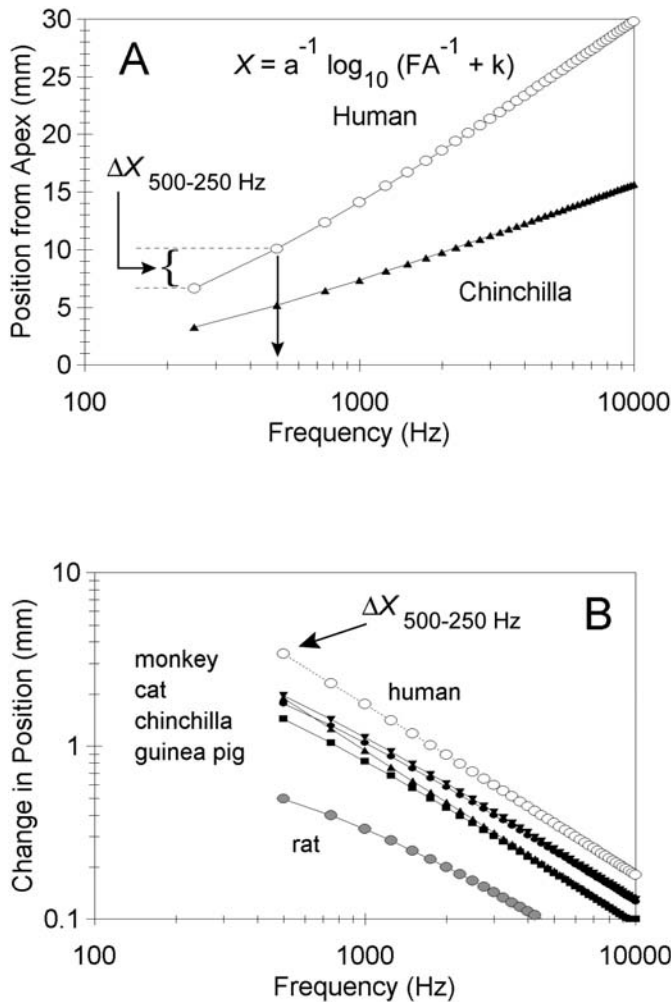


FIGURE 3.11. (A) Frequency-position functions for humans and common laboratory mammals based on the function derived by Greenwood (1961, 1990). The frequency-position equation is indicated. The symbols mark the locations of frequency components for a 250-Hz fundamental harmonic complex tone. *Open circles* show human function; *black triangles* show chinchilla function. Indicated is the difference in position between the 500-Hz and 250-Hz components ( $\Delta X_{500-250 \text{ Hz}}$ ; the *downward arrow* indicates that  $\Delta X_{500-250 \text{ Hz}}$  will be plotted at 500 Hz in Figure 11B. (B) Changes in position (i.e.,  $\Delta X$ ) between adjacent components of a harmonic tone complex having a 250-Hz F0 predicted for humans and common laboratory mammals. *Open circles* show human function. *Black squares* show guinea pig function; *black circles* show cat function; *black triangles* show chinchilla function; *black inverted triangles* show monkey function. *Gray circles* show rat function.

mammals than in humans. Consequently, the predicted number of unresolved components along the nonhuman mammalian cochlea should be greater than that along the human cochlea. This conclusion suggests that periodicity information in the stimulus envelope may have more importance for 'pitch' perception in animals than for human pitch perception, since periodicity information in the envelope is dominated by information in the unresolved components. A similar analysis to that described above carried out for birds using the modified Greenwood equation as derived by Fay (1992b) predicts that the number of unresolved components is also higher along the avian basilar membrane. Although there is some evidence that auditory filters in birds may be narrower than those in human listeners (see Dooling et al. 2000), suggesting that the smaller distance between harmonic components could be offset by the increase in frequency resolution, the greater sensitivity of birds to phase manipulations (as described in this chapter) may be an indication that there are indeed more unresolved components, and thus a greater contribution of the stimulus envelope to the perception. Although it can be concluded that the central mechanisms may be similar among animals and humans, experimenters should at least be aware of these kinds of potential differences in the peripheral representation of complex, periodic sounds among animals and humans.

## *6.2 Issue 2: Acoustic Environment and Listening Experience*

Normal-hearing human listeners live in an acoustically rich environment. From birth throughout adulthood, humans are exposed constantly to speech and music through everyday experience. Terhardt (1974) proposed that virtual pitch is acquired through a learning process, presumably related to the learning of speech sounds. Although this learning stage for virtual pitch has been supported by anecdotal evidence described by Divenyi (1979), it was pointed out that there is no way to test the learning aspects of Terhardt's model directly. In other words, it would essentially be impossible to find normal-hearing human subjects who have been raised in acoustically impoverished environments. How might the issue of the acoustic, listening environment relate to 'pitch' perception in animals? Some of the bird species used in the studies described earlier were collected from the wild, and therefore, presumably have experience with listening to bird songs. However, with a few exceptions, most animals used in laboratory experiments are raised and housed in environments that are impoverished acoustically. That is, most laboratory animals do not have the exposure to a rich acoustic environment in a way that matches the environments of normal-hearing human listeners. However, it would be relatively easy to compare 'pitch' perception in animals that have been raised in acoustically enriched and impoverished environments. The effects of listening environment and listening experience are issues that should be explored in future animal studies.

### 6.3 Issue 3: Establishing Neural Correlates for Perception

As mentioned in the introduction to this chapter, understanding behavior in animals is a necessary and important conceptual bridge between psychophysical studies in human listeners and neurophysiological experiments in animals. Thus, one goal of animal psychophysical experiments is to develop an appropriate “animal model” to then use to study the neurophysiological basis of human hearing. Many types of complex stimuli used in the behavioral experiments described in this chapter have also been used to study physiological responses of single units in the peripheral and central auditory systems of animals (see Winter, Chapter 4). To avoid comparisons across species, it is important to be able to account physiologically for the animal behavioral data, if those data are available. In other words, it is advantageous to relate cat physiological data to cat behavioral data, rather than to relate cat physiological data to human psychophysical data, for example. However, the latter comparisons are often unavoidable, because there simply exist fewer psychophysical and perceptual data from animals than from human listeners. It should also be noted that any physiological model of pitch perception will be based on data obtained from animal experiments, and thus the predictions of these models should also be compared to the appropriate animal behavioral data.

How do we relate the physiological responses to the behavior in order to establish a neural correlate? Certainly, the first step in this process is to establish that the physiological response of interest changes in a systematic manner, as does the percept or behavior. However, changes in physiological responses alone are insufficient to infer changes in behavioral sensitivity. The neural responses should be evaluated in the context of the behavioral task, using a physiological measure of sensitivity that is equivalent to behavioral performance. For example, Young and Barta (1986) used an analysis based on  $d'$  to examine the detection of a tone in noise in the auditory nerve of cats, whereas Relkin and Pelli (1987) used an analysis based on receiver operating characteristic curves to examine forward masking in the auditory nerve of chinchillas. Both of these approaches allowed the investigators to evaluate the optimal processing of average discharge rate from single auditory-nerve fibers and then make direct comparisons to existing psychophysical data. Because temporal responses (e.g., phase locking) have been implicated as being important in coding stimulus features related to pitch perception (see Winter, Chapter 4), the challenge for physiologists will be to develop approaches similar to those described by Young and Barta (1986) and Relkin and Pelli (1987) for evaluating optimal processing of temporal discharge patterns of single auditory units.

*Acknowledgments.* The preparation of this chapter was supported by NIH Grant P01 DC00293.

## References

- Amagai S, Dooling RJ, Shamma S, Kidd TL, Lohr B (1999) Detection of modulation in spectral envelopes and linear-rippled noises by budgerigars (*Melopsittacus undulatus*). *J Acoust Soc Am* 105:2029–2035.
- Au WWL, Pawloski JL (1989) Detection of noise with rippled spectra by the Atlantic bottlenose dolphin. *J Acoust Soc Am* 86:591–596.
- Bilsen FA, Ritsma RJ (1970) Some parameters influencing the perceptibility of pitch. *J Acoust Soc Am* 47:469–475.
- Blackwell HR, Schlosberg H. (1943) Octave generalization, pitch discrimination, and loudness thresholds in the white rat. *J Exp Psychol* 33:407–419.
- Braaten RF, Hulse SH (1991) A songbird, the European starling (*Sturnus vulgaris*), shows perceptual constancy for acoustic spectral structure. *J Comp Psychol* 105:222–231.
- Braaten RF, Hulse SH (1993) Perceptual organization of auditory temporal patterns in European starlings (*Sturnus vulgaris*). *Percept Psychophys* 54:567–578.
- Burns EM, Viemeister NF (1976) Nonspectral pitch. *J Acoust Soc Am* 60:863–869.
- Burns EM, Viemeister NF (1981) Played-again SAM: further observations on the pitch of amplitude-modulated noise. *J Acoust Soc Am* 70:1655–1660.
- Butler RA, Diamond IT, Neff WD (1957) Role of auditory cortex in discrimination of changes in frequency. *J Neurophysiol* 20:108–120.
- Capranica RR (1966) Vocal response of the bullfrog to natural and synthetic mating calls. *J Acoust Soc Am* 40:1131–1139.
- Chase AR (2001) Music discriminations by carp (*Cyprinus carpio*). *Anim Learn Behav* 29:336–353.
- Cranford JL, Igarashi M, Stramler JH (1976) Effect of auditory neocortex ablation on pitch perception in the cat. *J Neurophysiol* 39:143–152.
- Cynx J (1993) Auditory frequency generalization and a failure to find octave generalization in a songbird, the European starling (*Sturnus vulgaris*). *J Comp Psychol* 107:140–146.
- Cynx J (1995) Similarities in absolute and relative pitch perception in songbirds (starling and zebra finch) and a nonsongbird (pigeon). *J Comp Psychol* 109:261–267.
- Cynx J, Shapiro M (1986) Perception of missing fundamental by a species of songbird (*Sturnus vulgaris*). *J Comp Psychol* 100:356–360.
- Cynx J, Hulas SH, Polyzois S (1986) A psychophysical measure of pitch discrimination loss resulting from a frequency range constraint in European starlings (*Sturnus vulgaris*). *J Exp Psychol Anim Behav Proc* 12:394–402.
- Cynx J, Williams H, Nottebohm F (1990) Timbre discriminations in zebra finch (*Taeniopygia guttata*) song syllables. *J Comp Psychol* 104:303–308.
- D'Amato MR, Colombo M (1988) On tonal pattern perception in monkeys (*Cebus apella*). *Anim Learn Behav* 16:417–424.
- D'Amato MR, Salmon DP (1982) Tune discrimination in monkeys (*Cebus apella*) and in rats. *Anim Learn Behav* 10:126–134.
- D'Amato MR, Salmon DP (1984) Processing of complex auditory stimuli (tunes) by rats and monkeys (*Cebus apella*). *Anim Learn Behav* 12:184–194.
- Divenyi PL (1979) Is pitch a learned attribute of sounds? Two points in support of Terhardt's pitch theory. *J Acoust Soc Am* 66:1210–1213.
- Dooling RJ, Searcy MH (1981) Amplitude modulation thresholds for the parakeet (*Melopsittacus undulatus*). *J Comp Physiol* 143:383–388.

- Dooling RJ, Brown SD, Park TJ, Okanoya K, Soli SD (1987a) Perceptual organization of acoustic stimuli by budgerigars (*Melopsittacus undulatus*): I. Pure tones. *J Comp Psychol* 101:139–149.
- Dooling RJ, Park TJ, Brown SD, Okanoya K, Soli SD (1987b) Perceptual organization of acoustic stimuli by budgerigars (*Melopsittacus undulatus*): II. Vocal signals. *J Comp Psychol* 101:367–381.
- Dooling RJ, Lohr, B, Dent ML (2000) Hearing in birds and reptiles. In: Dooling RJ, Fay RR, Popper AN (eds), *Comparative Hearing: Birds and Reptiles*. New York: Springer-Verlag, pp. 308–359.
- Dooling RJ, Leek MR, Gleich O, Dent ML (2002) Auditory temporal resolution in birds: Discrimination of harmonic complexes. *J Acoust Soc Am* 112:748–759.
- Echteler SM, Fay RR, Popper AN (1994) Structure of the mammalian cochlea. In: Fay RR, Popper AN (eds), *Comparative Hearing: Mammals*. New York: Springer-Verlag, pp. 134–171.
- Fastl H, Weinberger, M (1981) Frequency discrimination for pure and complex tones. *Acustica* 49:77–78.
- Fay RR (1970) Auditory frequency generalization in the goldfish (*Carassius auratus*). *J Exp Anal Behav* 14:353–360.
- Fay RR (1972) Perception of amplitude-modulated auditory signals by the goldfish. *J Acoust Soc Am* 52:660–666.
- Fay RR (1982) Neural mechanisms of an auditory temporal discrimination by the goldfish. *J Comp Physiol* 147:201–216.
- Fay RR (1988) *Hearing in Vertebrates: A Psychophysics Databook*. Winnetka, IL: Hill-Fay Associates.
- Fay RR (1992a) Analytic listening by the goldfish. *Hear Res* 59:101–107.
- Fay RR (1992b) Structure and function in sound discrimination among vertebrates. In: Webster DB, Fay RR, Popper AN (eds), *The Evolutionary Biology of Hearing*. New York: Springer-Verlag, pp. 229–263.
- Fay RR (1994a) Comparative auditory research. In: Fay RR, Popper AN (eds), *Comparative Hearing: Mammals*. New York: Springer-Verlag, pp. 1–17.
- Fay RR (1994b) Perception of temporal acoustic patterns by the goldfish (*Carassius auratus*). *Hear Res* 76:158–172.
- Fay RR (1995) Perception of spectrally and temporally complex sounds by the goldfish (*Carassius auratus*). *Hear Res* 89:146–154.
- Fay RR (1998) Auditory stream segregation in goldfish (*Carassius auratus*). *Hear Res* 120:69–79.
- Fay RR (2000) Spectral contrasts underlying auditory stream segregation in goldfish (*Carassius auratus*). *J Assoc Res Otolaryngol* 1:120–128.
- Fay RR, Passow B (1982) Temporal discrimination in the goldfish. *J Acoust Soc Am* 72:753–760.
- Fay RR, Yost WA, Coombs S (1983) Psychophysics and neurophysiology of repetition noise processing in a vertebrate auditory system. *Hear Res* 12:31–55.
- Fay RR, Chronopoulos M, Patterson RD (1996) The sound of a sinusoid: perception and neural representations in the goldfish (*Carassius auratus*). *Audit Neurosci* 2:377–392.
- Flanagan JL, Saslow MG (1958) Pitch discrimination for synthetic vowels. *J Acoust Soc Am* 30:435–442.
- Formby C (1985) Differential sensitivity to tonal frequency and to the rate of amplitude

- modulation of broadband noise by normally hearing listeners. *J Acoust Soc Am* 78: 70–77.
- Gerhardt HC (1981) Mating call recognition in the barking treefrog (*Hyla gratiosa*): responses to synthetic calls and comparisons with the green treefrog (*Hyla cinerea*). *J Comp Physiol* 144:17–25.
- Green DM, Kidd Jr G (1983) Further studies of auditory profile analysis. *J Acoust Soc Am* 73:1250–1265.
- Greenwood DD (1961) Critical bandwidth and the frequency coordinates of the basilar membrane. *J Acoust Soc Am* 33:1344–1356.
- Greenwood DD (1990) A cochlear frequency-position function for several species—29 years later. *J Acoust Soc Am* 87:2592–2605.
- Guttman N (1963) Laws of behavior and facts of perception. In: Koch S. (ed), *Psychology: A Study of Science*, Vol. 5. New York: McGraw-Hill, pp. 114–178.
- Hainfeld CA, Boatright-Horowitz SL, Boatright-Horowitz SS, Simmons AM (1996) Discrimination of phase spectra in complex sounds by the bullfrog (*Rana catesbeiana*). *J Comp Physiol* 179:75–87.
- Heffner H, Whitfield IC (1976) Perception of the missing fundamental by cats. *J Acoust Soc Am* 59:915–919.
- Henning GB, Grosberg SL (1968) Effect of harmonic components on frequency discrimination. *J Acoust Soc Am* 44:1386–1389.
- Houstma AJM (1995) Pitch perception. In: Moore BCJ (ed), *Hearing. Handbook of Perception and Cognition*, 2nd ed: San Diego: Academic Press, pp. 267–295.
- Hulse SH (1995) The discrimination-transfer procedure for studying auditory perception and perceptual invariance in animals. In: Klump GM, Dooling RJ, Fay RR, Stebbins WC (eds), *Methods in Comparative Psychoacoustics*. Basel: Birkhauser Verlag, pp. 319–330.
- Hulse SH, Cynx J (1985) Relative pitch perception is constrained by absolute pitch in songbirds (*Mimus*, *Molothrus*, and *Sturnus*). *J Comp Psychol* 99:176–196.
- Hulse SH, Cynx J (1986) Interval and contour in serial pitch perception by a passerine bird, the European starling (*Sturnus vulgaris*). *J Comp Psychol* 100:215–228.
- Hulse SH, Cynx J, Humpal J (1984) Absolute and relative pitch discrimination in serial pitch perception by birds. *J Exp Psychol: Gen* 113:38–54.
- Hulse SH, Bernard DJ, Braaten RF (1995) Auditory discrimination of chord-based spectral structure by European starlings (*Sturnus vulgaris*). *J Exp Psychol: Gen* 124:409–423.
- Izumi A (2000) Japanese monkeys perceive sensory consonance of chords. *J Acoust Soc Am* 108:3073–3078.
- Izumi A (2001) Relative pitch perception in Japanese monkeys (*Macaca fuscata*). *J Comp Psychol* 115:127–131.
- Izumi A (2002) Auditory stream segregation in Japanese monkeys. *Cognition* 82:B113–B122.
- Jenkins HM, Harrison, RH (1960) Effect of discrimination training on auditory generalization. *J Exp Psychol* 59:246–253.
- Kieffe M, Kluender KR, Rhode WS (2002) Synthetic speech stimuli spectrally normalized for nonhuman cochlear dimensions. *Acoust Res Lett Online* 3:41–46.
- Leek MR, Summers V (2001) Pitch strength and pitch dominance of iterated rippled noises in hearing-impaired listeners. *J Acoust Soc Am* 109:2944–2954.
- Lohr B, Dooling RJ (1998) Detection of changes in timbre and harmonicity in complex

- sounds by zebra finches (*Taeniopygia guttata*) and budgerigars (*Melopsittacus undulatus*). *J Comp Psychol* 112:36–47.
- Long GR, Clark WW (1984) Detection of frequency and rate modulation by the chinchilla. *J Acoust Soc Am* 75:1184–1190.
- Lundeen C, Small Jr. AM (1984) The influence of temporal cues on the strength of periodicity pitches. *J Acoust Soc Am* 75:1578–1587.
- MacDougall-Shackleton SA, Hulse SH (1996) Concurrent absolute and relative pitch processing by European starlings (*Sturnus vulgaris*). *J Comp Psychol* 110:139–146.
- MacDougall-Shackleton SA, Hulse SH, Gentner TQ, White W (1998) Auditory scene analysis by European starlings (*Sturnus vulgaris*): perceptual segregation of tone sequences. *J Acoust Soc Am* 103:3581–3587.
- Malott RW, Malott MK (1970) Perception and stimulus generalization. In: Stebbins WC (ed), *Animal Psychophysics: The Design and Conduct of Sensory Experiments*. New York: Appleton-Century-Crofts, pp. 363–400.
- Mathes RC, Miller RL (1947) Phase effects in monaural perception. *J Acoust Soc Am* 19:780–797.
- Moody DB (1994) Detection and discrimination of amplitude-modulated signals by macaque monkeys. *J Acoust Soc Am* 95:3499–3510.
- Moody DB, LePrell CG, Niemiec AJ (1998) Monaural phase discrimination by macaque monkeys: use of multiple cues. *J Acoust Soc Am* 103:2618–2623.
- Moore BCJ (1993) Frequency analysis and pitch perception. In: Yost WA, Popper AN, Fay RR (eds), *Human Psychophysics*. New York: Springer-Verlag, New York, pp. 56–113.
- Moore BCJ, Glasberg BR (1988) Effects of the relative phase of the components on the pitch discrimination of complex tones by subjects with unilateral cochlear impairments. In: Duifhuis H, Horst JW, Wit HP (eds), *Basic Issues in Hearing*. San Diego: Academic Press, pp. 421–430.
- Moore BCJ, Peters RW (1992) Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity. *J Acoust Soc Am* 91:2881–2893.
- Moore BCJ, Glasberg BR, Shailer MJ (1984) Frequency and intensity difference limens for harmonics with complex tones. *J Acoust Soc Am* 75:550–561.
- Moore BCJ, Peters RW, Glasberg BR (1985) Thresholds for the detection of inharmonicity in complex tones. *J Acoust Soc Am* 77:1861–1867.
- Niemiec AJ, Yost WA, Shofner WP (1992) Behavioral measures of frequency selectivity in the chinchilla. *J Acoust Soc Am* 92:2636–2649.
- Ohm FW, Wetzel W, Wagner T, Rech A, Scheich H (1999) Bilateral ablation of auditory cortex in Mongolian gerbil affects discrimination of frequency modulated tones but not of pure tones. *Learning Memory* 6:347–362.
- Page SC, Hulse SH, Cynx J (1989) Relative pitch perception in the European starling (*Sturnus vulgaris*): further evidence for an elusive phenomenon. *J Exp Psychol: Anim Behav Proc* 15:137–146.
- Patterson RD (1994a) The sound of a sinusoid: spectral models. *J Acoust Soc Am* 96:1409–1418.
- Patterson RD (1994b) The sound of a sinusoid: time-interval models. *J Acoust Soc Am* 96:1419–1428.
- Poli M, Previde EP (1991) Discrimination of musical stimuli by rats (*Rattus norvegicus*). *Int J Comp Psychol* 5:7–18.



- Porter D, Neuringer A (1984) Music discriminations by pigeons. *J Exp Psychol Anim Behav Proc* 10:138–148.
- Recio A, Rhode WS, Kieft M, Kluender KR (2002) Responses to cochlear normalized speech stimuli in the auditory nerve of cat. *J Acoust Soc Am* 111:2213–2218.
- Relkin EM, Pelli DG (1987) Probe tone thresholds in the auditory nerve measured by two-interval forced-choice procedures. *J Acoust Soc Am* 82:1679–1691.
- Schulze H, Scheich H (1999) Discrimination learning of amplitude modulated tones in Mongolian gerbils. *Neurosci Lett* 261:13–16.
- Shera CA, Guinan Jr JJ, Oxenham AJ (2002) Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. *Proc Natl Acad Sci USA* 99:3318–3323.
- Shofner WP (2000) Comparison of frequency discrimination thresholds for complex and single tones in chinchillas. *Hear Res* 149:106–114.
- Shofner WP (2002) Perception of the periodicity strength of complex sounds by the chinchilla. *Hear Res* 173:69–81.
- Shofner WP, Selas G (2002) Pitch strength and Stevens' power law. *Percept Psychophys* 64:437–450.
- Shofner WP, Yost WA (1995) Discrimination of rippled-spectrum noise from flat-spectrum noise by chinchillas. *Audit Neurosci* 1:127–138.
- Shofner WP, Yost WA (1997) Discrimination of rippled-spectrum noise from flat-spectrum noise by chinchillas: evidence for a spectral dominance region. *Hear Res* 110:15–24.
- Simmons AM (1988) Selectivity for harmonic structure in complex sounds by the green treefrog (*Hyla cinerea*). *J Comp Physiol* 162:397–403.
- Simmons AM, Bean ME (2000) Perception of mistuned harmonics in complex sounds by the bullfrog (*Rana catesbeiana*). *J Comp Psychol* 114:167–173.
- Simmons AM, Buxbaum RC, Mirin MP (1993) Perception of complex sounds by the green treefrog, *Hyla cinerea*: envelope and fine-structure cues. *J Comp Physiol* 173:321–327.
- Simmons AM, Eastman KM, Simmons, JA. (2001) Autocorrelation model of periodicity coding in bullfrog auditory nerve fibers. *Acoust Res Lett Online* 2:1–6.
- Spiegel MF, Watson CS (1984) Performance on frequency-discrimination tasks by musicians and nonmusicians. *J Acoust Soc Am* 76:1690–1695.
- Stevens SS, Volkman J (1940) The relation of pitch to frequency: a revised scale. *Am J Psychol* 53:329–353.
- Symmes D (1966) Discrimination of intermittent noise by macaques following lesions of the temporal lobe. *Exp Neurol* 16:210–214.
- Terhardt E (1974) Pitch, consonance, and harmony. *J Acoust Soc Am* 55:1061–1069.
- Tomlinson RWW, Schwarz DWF (1988) Perception of the missing fundamental in non-human primates. *J Acoust Soc Am* 84:560–565.
- Tramo MJ, Shah GD, Braida LD (2002) Functional role of auditory cortex in frequency processing and pitch perception. *J Neurophysiol* 87:122–139.
- Whitfield IC (1980) Auditory cortex and the pitch of complex tones. *J Acoust Soc Am* 67:644–647.
- Wright AA, Rivera JJ, Hulse SH, Shyan M, Neiworth JJ (2000) Music perception and octave generalization in rhesus monkeys. *J Exp Psychol Gen* 129:291–307.
- Yost WA (1982) The dominance region and rippled noise pitch: a test of the peripheral weighting model. *J Acoust Soc Am* 72:416–425.

- Yost WA, Hill R (1978) Strength of the pitches associated with ripple noise. *J Acoust Soc Am* 64:485–492.
- Young ED, Barta PR (1986) Rate responses of auditory nerve fibers to tones in noise near masked threshold. *J Acoust Soc Am* 79:426–442.
- Zatorre RJ (1988) Pitch perception of complex tones and human temporal-lobe function. *J Acoust Soc Am* 84:566–572.

# The Neurophysiology of Pitch

IAN M. WINTER

## 1. Introduction

The representation of the pitch of a sound would appear to be a simple affair; the cochlea performs a spectral analysis of incoming sound and maps stimulus frequency onto place along the basilar membrane (BM). These mapped frequencies are then signaled to the brain via the auditory nerve (see Robles and Ruggero 2001 for a review). This tonotopic representation of a sound is often simulated using a computational model in which the membrane motion is represented by a bank of “auditory” filters (e.g., Patterson et al. 1995). The output of each filter is half-wave rectified and integrated to determine the activity level in that filter, and the set of levels is then plotted as a function of filter centre frequency (or cochlear place) to produce what is referred to as an “auditory spectrum” (see Fig. 2.3 in Plack and Oxenham, Chapter 2). This representation of tonotopic activity is often assumed to be the basis of pitch perception (e.g., Cohen et al. 1995).

It is also the case, however, that the inner hair cells (IHCs) transduce movement of the basilar membrane in-phase up to relatively high frequencies (e.g., approximately 5 kHz in the cat [*Felis catus*, Johnson 1980]; 3.5 kHz in the guinea pig [*Cavia porcellus*, Palmer and Russell 1986]). As a result, there is information about the timing of membrane peaks in each tonotopic channel. To make use of this information models have been developed that subject each frequency channel to autocorrelation (Slaney and Lyon 1990; Meddis and Hewitt 1991a), or some other form of temporal analysis (e.g., strobed temporal integration [Patterson et al. 1995]). The resulting two-dimensional representation (filter-center-frequency versus delay or time-interval) exhibits activity peaks across a range of channels at the period of pitch-producing sounds. Proponents of temporal models argue that it is this distribution of activity in these “auto-correlograms” that determines the perceived pitch (e.g., Meddis and Hewitt 1991a,b; Yost et al. 1996).

In this context this chapter reviews the evidence that pitch is encoded by place, timing, or a combination of the two by examining the correspondences

between neural patterns of activity at various stages along the auditory pathway, and the auditory percept of pitch. Strictly speaking, all the studies that are discussed in this chapter will be searching for a neural representation of the pitch of simple and complex sounds; much of the work has taken place using anesthetized preparations and thus one is forced to look only for representations and not a code. The concept of a neural code is reserved for the set of rules that relates behavior to neural activity (Eggermont 2001). Of necessity the neural activity has been recorded from nonhuman animals and this places an important constraint on the interpretation of any neural representation of pitch. The problems and successes of using animals to study the perception of pitch are discussed in detail by Shofner (Chapter 3). This chapter reflects the amount of information we have for the various parts of the auditory pathway; this information becomes increasingly sparse as we ascend from the auditory nerve to the auditory cortex. Although we arguably have most information about the mammalian cochlea, this chapter does not review the cochlea in any significant detail. For this information the interested reader is referred to reviews that can be found in a companion volume in the *Springer Handbook of Auditory Research*, Volume 8, *The Cochlea*. For a review of models of the processing of pure tones the reader is referred to the review by Delgutte (1996) (*Springer Handbook of Auditory Research*, Vol. 6: *Auditory Computation*). A review of models of the pitch of simple and complex sounds is provided by de Cheveigné (Chapter 6).

### 1.1 The Auditory Nerve

All information about the auditory stimulus has to be transferred from the cochlea to the brain via the VIIIth cranial nerve, the auditory nerve. In the cat the auditory nerve contains approximately 50,000 fibers, each terminating in the cochlear nucleus. The responses of single auditory nerve fibers are often described as homogeneous but they differ in important anatomical and physiological ways. Liberman (1978) divided auditory nerve fibers in the cat into three groups according to their spontaneous discharge rate (SR; spontaneous rate is defined as that discharge rate obtained in the absence of controlled acoustic stimulation). Approximately 60% of fibers have a high SR ( $> 18$  sp/s), are characterized by low thresholds and predominantly contact the pillar side of inner hair cells. In contrast, fibers with the lowest SR ( $< 0.5$  sp/s) form about 10% to 15% of the population of auditory nerve fibers, have the highest thresholds, and contact the modiolar side of the inner hair cell. Fibers with spontaneous rates between the other groups, known as medium-SR fibers, have intermediate thresholds and also contact the modiolar side of the IHC (Liberman 1978, 1982; Liberman and Oliver 1984). These differences in site of initiation are also reflected in their site of termination within the cochlear nucleus (e.g., Liberman, 1991, 1993). In the anteroventral cochlear nucleus the greatest number of auditory nerve fiber terminals belong to low-SR fibers and the small cell cap of the ventral cochlear nucleus (VCN) is almost exclusively innervated by

low and medium SR fibers. In contrast globular bushy cells (see Section 2.1) are innervated mainly by high-SR fibers while multipolar cells are innervated predominantly by low- and medium-SR fibers. The variation in threshold of the three fiber groups has important consequences for the dynamic range of individual fibers. High-SR fibers have the narrowest dynamic range (approximately 20 dB) while the low-SR, high-threshold fibers can have the largest dynamic range of any fiber group (Sachs and Abbas 1974; Winter et al. 1990). The relationship between dynamic range and fiber threshold was first proposed by Sachs and Abbas (1974), who hypothesized that the nonlinear growth of the basilar membrane motion as a function of sound level, combined with a saturating nonlinearity at the IHC/synapse, could account for the different dynamic ranges. This theory received experimental support from the study of Yates et al. (1990) looking at the responses of single auditory nerve fibers in the guinea pig. They showed that it was possible to change the shape of the rate-level function by changing the threshold of the auditory nerve fiber. This threshold change was implemented by forward masking the response of the auditory nerve fiber. It is reasonable, given the different sites of origin and termination for the three SR groups and their differing physiology, to suggest that parallel processing of information about a sound begins at the level of the auditory nerve.

#### 1.1.1 Rate-Place Representations of the Pitch of Single Tones

Arguably, the simplest mechanism for the representation of the pitch of a pure tone suggests that the pitch is directly correlated with the “place” of maximum discharge rate in the peripheral auditory system. This “labeled-line” mechanism is more commonly referred to as a rate-place code and is often measured by recording of the responses of a large number ( $>> 100$ ) of auditory nerve fibers to the same stimulus, from the same animal—a who’s listening paradigm (Pfeiffer and Kim 1975; Kim and Molnar 1979; Evans and Palmer 1980; Shofner and Sachs 1986; Kim and Parham 1990; Kim et al. 1990). The picture that has emerged is that at reasonably low sound levels a plot of average discharge rate versus CF (a rate-place profile) for high SR fibers shows a peak around the frequency position of the pure tone. At high sound levels a peak remains only in the discharge rate of low spontaneous rate auditory nerve fibers.

Kim and Parham (1990) examined the responses of a population of auditory nerve fibers to a 5-kHz tone, where the use of temporal information is less obvious (see Section 1.1.2). They examined the statistical properties of a population of cat auditory nerve fiber discharge patterns to produce a measure of the spatial discrimination of spike discharge pattern along the length of the cochlea. From this analysis they concluded that at low sound levels (30 dB SPL) the discriminability was greatest for fibers with high SR; at higher sound levels (50 to 70 dB SPL) the discriminability was greater for low SR fibers. Therefore there was more than sufficient information in a rate-place code to support behavioral frequency discrimination if the cat was able to optimally combine information from the different SR groups. The potential importance

of the low-SR population in encoding the pitch of pure tones at high levels has also been demonstrated by Shofner and Sachs (1986), who found a clear peak at the 1.5-kHz place in a population of fibers with low SR at moderately high sound levels (86 dB SPL). This result, combined with that of Kim and Parham (1990), indicates that there is potential information in the mean rate discharges of low SR auditory nerve fibers at high sound levels for both high- and low-frequency regions of the cochlea. A similar analysis was carried out by Kim et al. (1990a) looking at the responses of a population of auditory nerve fibers to a 1-kHz tone. In this study they demonstrated that the discharge statistics of low-SR fibers was particularly well suited to represent the frequency position and level of the 1-kHz tone in a rate-place profile. However, they also noted a small shift in the frequency position of the peak to more apical regions at 70 dB SPL relative to that seen in the 30 dB SPL rate-place profile. This was attributed to nonlinearities in cochlear mechanics and may be related to the small shift in the perception of F0 with increases in sound level. Further studies are needed to determine if the direction of the frequency shift in the population of nerve fibers is frequency dependent, as is observed in the psychophysics. Kim et al. (1990a) argued that the reason for the success of the low-SR fibers was the reduction in the variance of their discharge with increases in sound level but a similar reduction in spike discharge variance has not been reported by others (e.g., Young and Barta 1986; Delgutte 1987). If low-SR auditory nerve fibers are involved in the representation of spectral peaks in a rate-place profile, then a more central nucleus must be capable of combining the information from the different fiber groups. One theory suggests that cells in the cochlear nucleus are able to respond to high-SR fibers at low sound levels and switch their attention to the unsaturated, low-SR fibers at high intensities (Delgutte 1982; Winslow and Sachs 1988; see Section 2.2.1). The limited dynamic range of individual auditory nerve fibers and the necessity of combining information about sound level across the different SR groups in, as yet, unproven theories, has led others to explore alternate means of encoding frequency at high sound levels. For instance the phase-opponency model uses the relative timing differences across auditory nerve fibers with different CFs (Carney 1994; Carney et al. 2002). These timing differences are then hypothesized to be extracted at the level of the cochlear nucleus by coincidence detectors (see Section 2.1).

### 1.1.2 Temporal Representations of the Pitch of Single Tones

The nature of the transduction process in the cochlea ensures that primary auditory nerve fibers discharge at preferred phases of the stimulating waveform. This nonrandom discharge pattern is referred to as phase locking and is evident in the discharges of neurons at all levels of the auditory pathway, from the auditory nerve to the auditory cortex (Johnson 1980; Wallace et al. 2002). Examples of phase-locked discharges are shown in Figure 4.1A for a low best-frequency (BF) unit recorded from the cochlear nucleus in response to a 300 Hz tone burst. Phase locking has been quantified using a variety of related

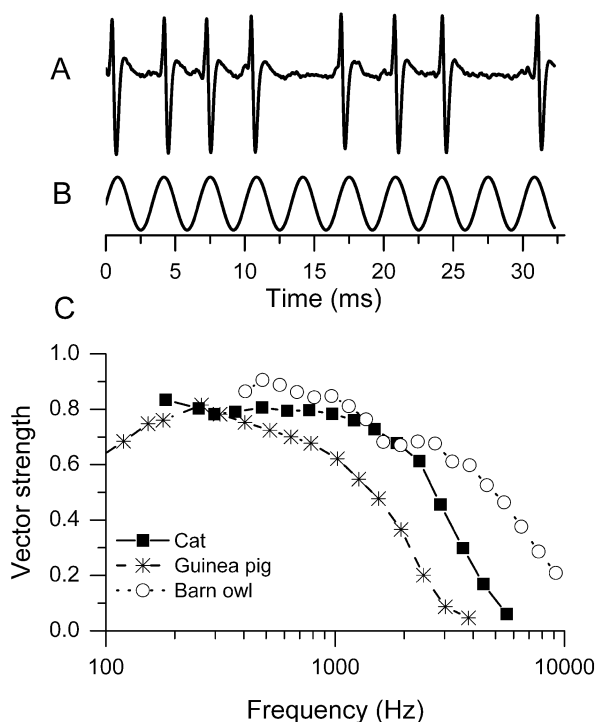


FIGURE 4.1. Phase locking in the auditory pathway. The neuron was a chopper unit ( $BF = 0.9$  kHz) in the cochlear nucleus of the guinea pig responding to a 300 Hz tone. *Top trace (A)* is the extracellular spike waveform. Note that you do not get an action potential at every period of the waveform. *Bottom trace (B)* is the stimulus waveform. *(C)* Vector strength (a measure of phase locking) as a function of frequency for three species commonly used in the study of the auditory system. Note the substantial differences in the upper frequency limit: approximately 3.5 kHz in the guinea pig (Palmer and Russell 1986); approximately 5 kHz in the cat (Johnson 1980), and  $> 10$  kHz in the barn owl (Koppl 1997). Data kindly provided by Christine Koppl.

measures (e.g., vector strength, Goldberg and Brown 1969; synchronization index, Johnson 1980; periodicity strength, Kim et al. 1986) and the magnitude of phase locking declines with increasing frequency. However, the corner frequency and slope of this decline appears to be species dependent. In the cat the corner frequency is approximately 2.5 kHz and the synchronization index (SI) drops to below 0.1 around 5 kHz. In the guinea pig the corner frequency is as low as 1.1 kHz and the SI is less than 0.1 around 3 kHz (Weiss and Rose 1988). The barn owl (*Tyto albus*) is the current world record holder with an SI of 0.2 even at 10 kHz (see Fig. 4.1C). The decline in phase locking with increases in frequency has been attributed to the low-pass filtering found in the inner hair cell and its synapse (Palmer and Russell 1986; Weiss and Rose 1988).

Although many models of pitch processing rely on temporal discharge patterns (see de Cheveigné, Chapter 6), it is still unknown how this temporal information could be used by the auditory system in the processing of pitch or whether it is simply an epiphenomenon of the transduction process. Furthermore, the upper limit of phase locking decreases as one ascends the auditory pathway so that by the time information reaches the inferior colliculus the upper limit has fallen to approximately 600 Hz and by the time the information reaches the cortex it has fallen still further to approximately 250 Hz (Wallace et al. 2002). Obviously, if this temporal information is used by the auditory system, then it has to be recoded at a fairly early stage of the auditory pathway.

Using a model based closely on the responses of the mammalian cochlea and auditory nerve to single tones, Heinz et al. (2001) have argued that frequency discrimination is more likely to be based on temporal than rate-place information. At 1 kHz information contained in temporal discharges was an order of magnitude better than that obtained by a rate-place mechanism. In addition, the performance of a group of high SR fibers using a mean rate code was constant as a function of increasing frequency. In contrast, in keeping with human psychophysical performance (see Fig. 2.1, in Plack and Oxenham, Chapter 2, and Fig. 6.8, in de Cheveigné, Chapter 6) predicted frequency discrimination performance based on the temporal discharge properties decreased with increasing frequency up to around 10 kHz. This is remarkable given that Heinz et al. (2001) used the synchronization data of the cat. Clearly we, or rather the cat, could potentially use the very low synchronization values present at frequencies above 5 kHz. For the rate-place code to work one would have to postulate that there was a central deficit in the processing of mean-rate information above about 2 kHz. Proponents of temporal theories also argue that we cannot perceive the pitch of musical instruments above 4 to 5 kHz—again very close to the upper limit of phase locking in cats (Semal and Demany 1990) and that a weak pitch can be perceived when listening to the sinusoidally amplitude-modulated white noise (SAM noise). The long-term spectrum of SAM noise is flat and therefore the information about the pitch is likely to be related to the modulated waveform. However, the change in the pitch of pure tones with increases in sound level (see Plack and Oxenham, Chapter 2) cannot readily be explained by temporal theories but this is as much a problem for rate-place codes, and we currently have no satisfactory explanation for why the pitch of pure tones varies with sound level. Although the limited dynamic range of individual auditory nerve fibers is a problem for rate-place theories, this could be overcome by the differential weighting of the three SR fiber types, with the greatest weight given to the low-SR, high-threshold fibers at high sound levels (e.g., Delgutte 1987). Alternatively, cells in the cochlear nucleus could make use of the phase information present in the auditory nerve fibers to extract information about stimulus level. However, whatever the mechanism, it is clear the most important issue now is how information in the auditory nerve is combined and/or extracted at the level of the cochlear nucleus.



## *1.2 The Representation of the F0 of Complex Sounds*

At first glance the perception of the fundamental frequency (F0) of complex sounds appears to provide severe difficulties for rate-place theories of pitch. For many complex sounds the F0 does not correspond to the position of maximum excitation along the basilar membrane. For instance, consider a sound consisting of a set of harmonics with frequencies 200, 400, 600, and so on. This sound has a low pitch of 200-Hz, even when it is filtered to remove the 200 Hz component. The low pitch associated with a group of high harmonics has been called the residue pitch (Schouten 1940). Importantly, this low, residue, pitch persists even in the presence of low-pass masking noise which is assumed to mask the presence of cochlear distortion (see McAlpine 2004). There are essentially two rival explanations to explain the perception of the residue: pattern recognition models and temporal models. The pattern recognition models represent the frequency of individual harmonics of a complex sound and then estimate the pitch from these resolved harmonics. In contrast, the temporal models usually require the interaction of two harmonics. Both theories suffer from a lack of physiological evidence although the problem is arguably greater for the pattern recognition model, which requires the comparison of an input spectrum with an internally stored spectral template. Just how such an internal template may arise has been modeled by Shamma and Klein (2000), who also attempted to identify possible physiological and anatomical correlates of their model; however, experimental confirmation awaits.

For pattern recognition models it is therefore important to ascertain the ability of auditory nerve fibers to resolve individual harmonics of complex sounds. It is now widely believed that frequency resolution is determined by the shape of the auditory filters and in turn the bandwidth of these filters is determined by the vibration patterns of the BM (see Robles and Ruggero 2001 for a review). The vibration pattern of the BM is reasonably well recapitulated in the discharges of auditory nerve fibers, at least for high frequencies. Of course we do not have a direct indication of the width of auditory nerve fiber filters in humans. However, in a study comparing the bandwidths of tuning curves obtained using the same methods used to determine filter shape in humans, Evans (2001) has shown that behavioral bandwidths in the guinea pig and the bandwidths of single auditory nerve fibers in the same species show substantial overlap (Fig. 4.2). This is an important result, as it suggests that we can expect human frequency resolution to be reasonably accurately predicted by the psychophysical measurement of filter shapes. This result has now been repeated in humans but with the important difference that the “physiological” data agreed with psychophysical estimates of tuning obtained using forward masking (Shera et al. 2002, Oxenham and Shera 2003). Forward masking is used to avoid suppressive interactions between the masker and signal and is thus thought to give a more accurate estimate of auditory filter bandwidth. This, of course, presents us with a problem in interpreting the guinea pig behavioral data, which were obtained

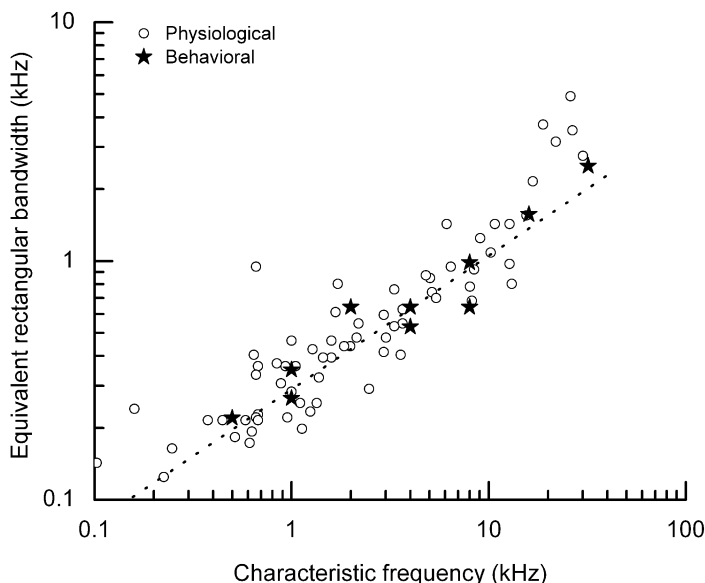


FIGURE 4.2. The relationship between psychophysical and physiological bandwidths in the guinea pig (replotted from Evans 2001). *Open circles* represent single auditory nerve fibers; *filled stars* represent measurements of psychophysical (behavioral) bandwidths derived using either a bandstop noise technique or comb-filtered noise. The *dashed line* is the derived relationship between characteristic frequency and equivalent rectangular bandwidth ( $ERB = 0.29CF^{0.56}$  where CF is in kHz).

using simultaneous masking; either the physiological recordings in the guinea pig are an overestimate of auditory nerve fiber bandwidth or the amount of suppression is less in the guinea pig than in the human. It may also be due to the relatively poor frequency tuning in the guinea pig, which is generally worse than in humans by a factor of 2 or more. These issues have yet to be fully resolved.

Responses of auditory nerve fibers to harmonic complex sounds with nearly flat spectra have shown that, at low sound levels, the temporal responses generally reflect components at or near fiber CF (Evans 1981; Horst et al. 1985 1986, 1990). The picture at higher sound levels is less predictable: the bandwidth of the response spectrum generally increases while the frequency of the component that dominates the response decreases. This is entirely consistent with level dependent nonlinearities at the level of the cochlea (Robles and Ruggero 2001). In a study examining the effect of resolvability of the encoding of F0 in the discharges of single auditory nerve fibers in the cat, Cedolin and Delgutte (2005) have shown that the lowest F0 whose harmonics could be resolved in the mean rate output of an auditory nerve fiber increased with increases in CF, consistent with the progressive sharpening of the cochlear filters with

increasing CF. They also found that F0s in the range of human voices were not resolved by single auditory nerve fibers in the cat and that rate-place profiles were best for F0s above 400 Hz. However, F0s up to 1300 Hz were represented in pooled interspike interval distributions of auditory nerve fibers.

In a who's listening experiment using the consonant-vowel syllable /da/, Miller and Sachs (1984) showed that auditory nerve fibers with CFs that fell within spectral dips of the stimulus had a strong response to the F0. This is in contrast to the responses of fibers whose CFs fell near a formant frequency, where the responses were dominated by the formant frequency. A similar result was found by Delgutte and Kiang (1984) when looking at the responses of single auditory nerve fibers in the cat to steady-state vowels. Single fibers with CFs between the first two formant frequencies and above the second formant show broadband responses along with deep envelope modulation at the F0. The determining factor in whether a fiber responds to the F0 envelope is whether or not its response is dominated by a single large-stimulus component. Miller and Sachs (1984) also found clear, harmonically related peaks in a temporal-place representation and these could be used by the auditory system to signal the pitch. Using a Cepstral analysis (a Fourier transform of the logarithm of the magnitude spectrum, or in this case the temporal-place representation), they demonstrated a strong pitch-related peak. Interestingly the Cepstral analysis was relatively undisturbed by background noise but the response of fibers with CFs between formant peaks (i.e., those showing a strong response to the stimulus envelope) showed a large reduction in response to the F0 in the presence of background noise. Thus F0 can be represented by peaks in the temporal responses at harmonic places in the population of auditory nerve fibers. This representation is very similar to the one modeled by Srulovicz and Goldstein (1983) and generally supports pattern recognition models of F0 encoding. Of course, it does not provide a biological mechanism for the templates needed to extract this harmonic structure.

An alternative to the temporal-place mechanism is an analysis based on the predominant interspike intervals present in populations of auditory nerve fibers. However, for complex sounds the use of first order interspike intervals has proven to be level dependent. One way to overcome this problem is the processing of higher-order interspike intervals, an operation equivalent to an autocorrelation of the spike train (Shofner 1991, 1999; Cariani and Delgutte 1996a, b). A stimulus periodicity represented in first-order interspike intervals at low stimulus levels may be preserved in higher-order interspike intervals for higher sound levels. This was confirmed experimentally by Cariani and Delgutte (1996a,b), who found that a neural correlate of pitch in the cat auditory nerve is well preserved in an all-order interspike interval analysis, whereas a first-order analysis was susceptible to changes in sound level. The response of a population of auditory nerve fibers to a single-formant vowel with an F0 of 80 Hz shows that as stimulus level increases so does the position of the largest peak in the first-order interspike interval histogram. At both 40 and 80 dB the largest peaks were at intervals much shorter than the reciprocal of the F0. In contrast, the

position of the peak in the all-order interspike interval distribution remained unchanged over the same range (40 dB) of sound levels (Fig. 4.3). These population interval distributions are analogous to the summary autocorrelation often produced in temporal models of pitch perception. The hypothesis that the pitch of sounds is represented by the largest peak in the population all-order interspike intervals (the predominant interval hypothesis) was tested further by Cariani and

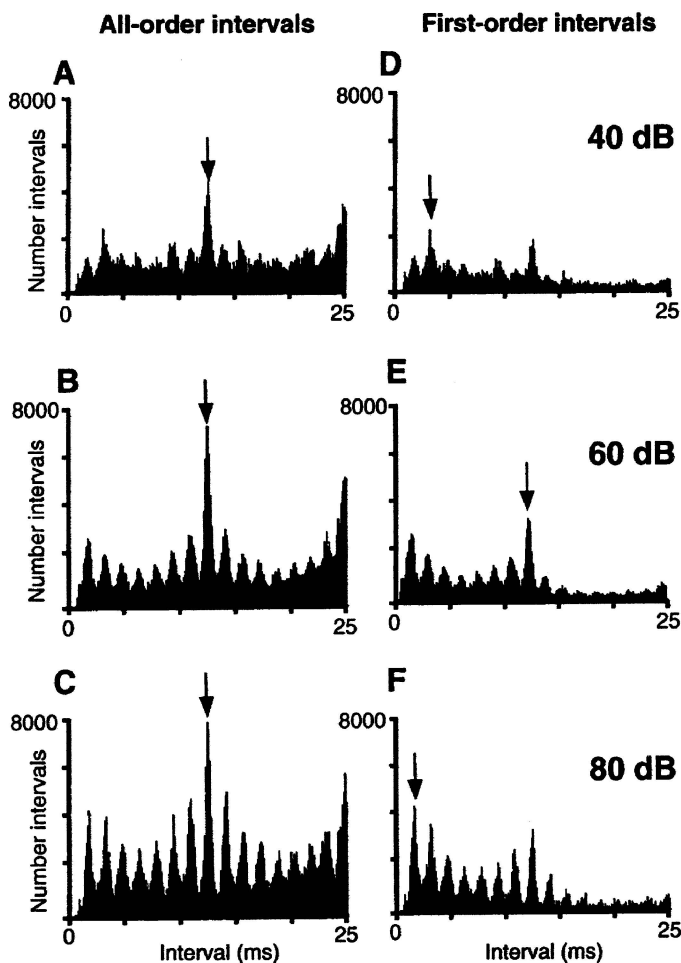


FIGURE 4.3. All-order interspike intervals are more level independent than first-order interspike intervals. This is demonstrated by looking at the distribution of all-order and first-order interspike intervals in a population of auditory nerve fibers of the cat in response to the pitch (80 Hz) of a single-formant vowel (Cariani and Delgutte 1996a). Note the change in the position of the most prominent interval (indicated by arrows) in the first-order representation as sound level is increased over a 40-dB range. In contrast the most prominent interval, 12.5 ms, is unchanged in the all-order representation.

Delgutte (1996a,b) by using stimuli that differed markedly in their power spectra but nevertheless evoked the same pitch. Stimuli as diverse as pure tones, amplitude-modulated tones, click trains, and amplitude-modulated noise all showed major interval peaks at the pitch period in the population interval distributions. In most cases sounds evoking the strongest pitches—pure tones and AM tones—produced population interval distributions with higher mean-to-peak ratios in comparison with stimuli that evoke a weak pitch (e.g., amplitude-modulated noise). Paradoxically, however, pure tones did not produce the highest mean-to-peak ratio.

The representation of the pitch of complex sounds by a mean rate code is more problematic. For instance, the recording of neurons with CFs equal to the low pitch of complex sounds is relatively rare and often inferences have to be made based on the responses of relatively high CFs and high pitches. Whether they translate to very low frequencies ( $< 300$  Hz) is speculative. Perhaps the best information we have of the encoding of spectral peaks in complex sounds comes from studies of steady-state vowels (Young and Sachs 1979; Delgutte and Kiang 1984; Palmer et al. 1986; May et al. 1998). In these studies, at relatively low sound levels, a clear representation of formant peaks can be found in a profile of mean discharge rate as a function of auditory nerve fiber CF. As sound level increases, however, the formant peaks become less clear. This is largely due to rate-saturation and the broadening of the auditory-nerve fiber filters at the higher stimulus levels. A representation of the formant peaks was still found if only fibers with low SR were analyzed (Sachs and Young 1979). A computational model, based on the distribution of the different types of SR fiber has shown that, in quiet, a good representation of not only the formant peaks but also the low harmonics of the steady-state vowel /e/ can be demonstrated in a rate-place profile (Delgutte 1996). However, the representation of formant peaks in the presence of background noise presents more of a challenge for rate-based codes.

### *1.3 The Representation of $F_0$ in the Presence of Competing Sounds*

In the presence of background noise even fibers with low SRs do not appear to give a good representation of the formant peaks of the steady-state vowel /e/. This is a potentially fatal blow to rate-place codes, however, it is possible that other mechanisms may enable a rate-place code to exist at disadvantageous S/N ratios. May and Sachs (1992) have shown that the dynamic range of single units in the ventral cochlear nucleus of the awake cat, to tones in background noise, is greater than the dynamic ranges found in anesthetized cats. They interpreted these results as suggesting that there was an active olivocochlear system in the awake animals that was helping to preserve the dynamic range of single units. Winslow and Sachs (1988) had previously demonstrated a similar decompression of the dynamic range in background noise following electrical stimulation of the olivocochlear bundle in single auditory nerve fibers of the cat.

Furthermore, May et al. (1996) have now shown that format peaks may be preserved in a rate-place code at high sound levels and in the presence of background noise when analyzing the discharges of low-SR fibers using statistical methods. Geisler and Silkes (1991) have shown that temporal discharges of low-SR fibers are also much better than high-SR fibers at representing the F0 of single vowels and the syllable murmur “m” in background noise even when the level of the noise was at the same level as the syllable, that is, at 0 dB signal-to-noise ratio. This result confirmed earlier studies by Miller and Sachs (1984), who found that the encoding of the F0 of noise-embedded syllables was less affected by noise for low-SR fibers, and this result further emphasizes the importance of low-SR fibers in the representation of F0 in the auditory nerve.

The presence of a competing voice presents the auditory system with an even harder task; voices share many spectral and temporal characteristics making the use of simple filtering ineffective. Double vowels, with a common F0, evoke the percept of a single talker producing a dominant vowel whose phonetic quality is colored by the impression of a second vowel. When a difference in F0 is introduced, accuracy of identification improves by as much as 20% at one semitone difference. However, in many cases human listeners can identify both members of a pair of vowels presented simultaneously, even when they share the same F0. When the difference in F0 is large enough to lead to improved discrimination performance the perception also changes. At larger F0 differences, listeners generally hear two voices rather than one, producing different vowels with different pitches. This indicates that the listener has established the presence of two F0s and correctly associated the formant-related peaks with the F0 from which they derive. Recording from single auditory nerve fibers, Palmer (1990) showed that the two F0s of a double vowel were visible in the modulation of the discharge of auditory-nerve fibers in frequency regions where individual harmonics were not resolved or where the discharge was not strongly dominated by a single strong component. This occurred in different frequency regions for the two F0s. The F0s of the double vowels could also be identified from the distribution of synchronized discharges across the population of nerve fibers or from computations based on intervals between discharges. Modeling studies (de Cheveigné, 1993) have shown that the F0s from two simultaneous harmonic stimuli can be extracted from the waveforms at the output of the auditory filter bank models. However, the situation is less clear in the neural data of Palmer (1990). In response to a double vowel stimulus with F0s of 100 and 125 Hz, a summary autocorrelation applied to the data of Palmer (1990) shows the largest peak is at 10 ms but the second largest peak, at 7.34 ms, is not at the second F0 (8 ms). For this set of data at least, a summary autocorrelogram is not an adequate representation of the two pitches of the double vowels.

#### *1.4 The Representation of the F0 of Click Trains*

Simple autocorrelation models of pitch perception have also been challenged with the use of click train stimuli. Although these stimuli have been described

as the type of sounds one would be forced to listen to in the fifth level of Hell in Dante's *Inferno* (Darwin, personal communication), they nevertheless provide a good test of mechanisms of temporal pitch. A simple interpretation of the autocorrelation model of pitch perception would predict that stimuli with the same first peak in their waveform autocorrelation would have the same pitch. That this is not true was demonstrated by Kaernbach and Demany (1998) using click trains with either first-order periodicity (regular intervals between successive peaks or higher-order periodicity (regular intervals between nonsuccessive clicks)). They described two types of click train with a single peak in the waveform autocorrelation. The first stimulus contained a regular interval followed by two random intervals and was called KXX. The second stimulus contained a regular interval formed by the addition of two random intervals followed by a single random interval, called ABX. A simple autocorrelation of the waveforms would predict equal pitch strength for the two stimuli. However, KXX was easier to discriminate from random click trains than ABX. Kaernbach and Demany (1998) interpreted this result as evidence against the use of autocorrelation and for the importance of first-order ISIs in the encoding of pitch. However, Pressnitzer et al. (2001) demonstrated that this result could be predicted by either a first-order or all-order representation if the autocorrelation analysis was not carried out on the stimulus but rather on the output of a model of the peripheral auditory system. Furthermore, in a modification of the original KXX stimulus, Pressnitzer and colleagues demonstrated that stimuli with the same first peak in the waveform autocorrelation could have different subjective pitches when passed through a model of the auditory periphery. However, the magnitude of the perceptual pitch shift between KXX and ABX (see Fig. 4.4A for the stimuli) was much smaller than the shift in either the first-order or all-order interspike interval distributions from a simulated auditory nerve fiber or a population of single units in the ventral cochlear nucleus of the guinea pig (Fig. 4.4B). This suggests that a weighting function must be applied to either the first-order or all-order representation for these distributions to represent the pitch of these stimuli (Pressnitzer et al. 2004). A similar challenge to the autocorrelation model has been provided by Carlyon et al. (2002), who have looked at the perception of click train sequences that were bandpassed between 3.5 and 5.3 kHz. These click trains had a sequence of 4- and 6-ms intervals. A first-order interval interpretation would suggest that the 4- and or 6-ms pitch should predominate. If an all-order analysis occurred then the pitch should be heard as 10ms. In fact neither pitch resulted but rather a pitch at 5.7ms. The authors argued that this could be explained with a weighted first-order interpulse interval interpretation; *longer* intervals would be given a stronger weight. Intriguingly, the physiological results of Pressnitzer et al. (2004) suggest that *shorter* intervals should be given more weight. Perhaps the most important conclusion to be extracted from the use of these stimuli is that simple first-order or all-order representations are not adequate to explain these results and that other transformations or weightings or even alternative ways of analyzing spike trains need to be considered.

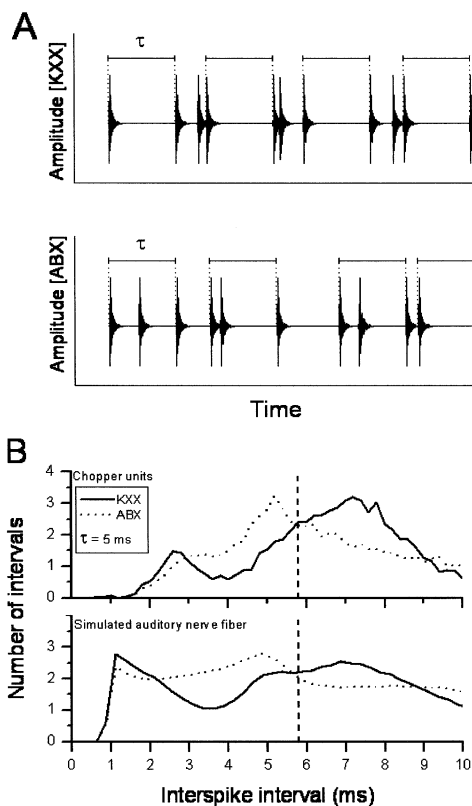


FIGURE 4.4. (A) Two high-pass filtered click train stimuli (ABX and KXX) with identical peaks in their autocorrelation. The regular interval is denoted as  $\tau$ , in this case 5 ms, both stimuli have the same average rate of clicks. (The two stimuli have the same peak in their autocorrelation function.) The peak is found at the interval,  $\tau$ . (B) First-order interspike interval histograms from a population of chopper units in the cochlear nucleus and a simulated auditory nerve fiber. The shift in activity below the regular interval for KXX is visible in first-order (FOIH—upper row) interspike interval histograms. In this case, large values are concentrated near the delay (5 ms) for the ABX stimuli, whereas there is a second peak, shifted toward longer delays in the KXX case. This is consistent in direction, but larger in magnitude, than the perceptual pitch shift which was at 5.8 ms (dotted vertical line).

## 2. The Cochlear Nucleus

As it is possible for the F0 of a variety of stimuli to be signaled to the brain using either rate-place or temporal information, the responses of cells in the cochlear nucleus (the termination site of all primary auditory-nerve fibers) may prove pivotal in deciding which codes have most utility in their progression to higher centers. On entering the cochlear nucleus the auditory nerve fibers bifurcate, sending one branch anteriorly to the anteroventral cochlear nucleus (AVCN) and the other branch dorsally, to the posteroventral (PVCN) and dorsal cochlear nucleus (DCN). The tonotopic organization of the cochlea is preserved independently in the three divisions of the cochlear nucleus (Rose et al. 1959). This preservation of tonotopicity has often been used to argue for the primacy of spectral theories of pitch extraction. However, the importance of this preservation is unclear; for instance, the tonotopic representation of a single tone in primary auditory cortex is strongly level dependent (see Section 4.1).

In comparison to its input from the auditory nerve the responses of single



units in the cochlear nucleus are more heterogeneous. Several cell types have been identified, both anatomically and physiologically, and it is reasonable to assume that each different cell type performs a different signal processing task. This is exemplified in Figure 4.5 which shows the responses of single units in the ventral cochlear nucleus to the steady-state vowel / $\epsilon$ / (Kim and Leonard 1988). The waveform of the vowel is shown at the top of the figure and has a

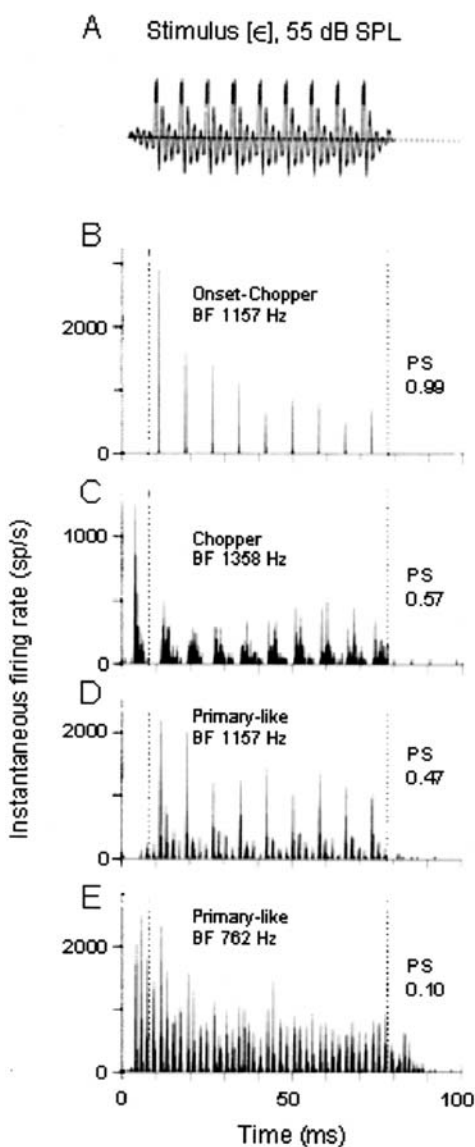


FIGURE 4.5. Responses of four cochlear nucleus units: onset (A), chopper (B), and primary-like (C) and (D) to the steady-state vowel / $\epsilon$ / (E). The vowel FO was 128 Hz. F1 = 512 Hz and F2 = 1792 Hz. The periodicity strength (PS) or temporal precision to the FO, shown inset of each poststimulus time histogram, is calculated using spikes that occur between the two dotted lines. Note that the best frequencies for the first three units, A–C, fell between the formant peaks. The unit classified as an onset-chopper has the highest PS while the other unit types have lower PS values indicating a broader spread of spike times in each period of the vowel. Data redrawn from Kim and Leonard (1988).

repetition period of approximately 7.8 ms ( $F_0 = 128$  Hz). Responses of four single units in the cochlear nucleus are shown as poststimulus time histograms. The response shown in the second row is from an onset-chopper unit (see Section 2.3). This unit responds with discharges locked to the repetition period of the vowel. The unit in the next row is from a chopper (see Section 2.2) and again one can see a clear preference in the temporal discharge characteristics to the repetition period; however, in this case the unit tends to fire more throughout the repetition period. Immediately below this is a response from a primary-like unit which shows the weakest representation of the repetition period but is also able to preserve some of the fine time structure of the vowel. It should be noted that the BFs of these units all fell between formant peaks, and in keeping with their input from auditory nerve fibers, they respond strongly to the  $F_0$ . This is probably due to the modulation in their input, that is, no single harmonic is able to dominate the response. In contrast, when one looks at the response of a primary-like unit with a BF near the first formant frequency (fifth row) it is dominated more by the first formant frequency (discharges are phase-locked to this frequency) than by the  $F_0$ . These results show a transformation of the responses observed in the auditory nerve. It should be borne in mind that unlike the population response profiles produced in the auditory nerve, it is extremely difficult to produce a large population of units in the central auditory pathway from one animal, and therefore population analyses are rare in studies on the cochlear nucleus.

## 2.1 *Primary-Like Units*

As their name implies, primary-like units have a characteristic temporal adaptation pattern that is very similar to that observed in their auditory nerve fiber input. Figure 4.6A shows a typical example of a temporal adaptation pattern for a primary-like unit plotted as a poststimulus time histogram (PSTH). The primary-like units are recorded from bushy cells in the anteroventral cochlear nucleus and are contacted by one or several large end bulb of Held-type synapses. This property enables the primary-like population to most faithfully preserve the temporal information present in the auditory nerve and it is often assumed that the primary-like units represent the best opportunity of getting precise temporal information to higher levels of the auditory pathway. In response to complex sounds primary-like units seem to behave in a fashion similar to their auditory nerve fiber input (e.g., Blackburn and Sachs 1990; Winter and Palmer 1990b). Given the similarity in the temporal responses of primary-like units and auditory nerve fibers, it is usually assumed that primary-like units must be conveying temporal information about the  $F_0$  of both simple and complex sounds to higher levels in the auditory pathway. In fact the precision of phase locking in some primary-like units exceeds that of the auditory nerve for low frequencies (Joris et al. 1994). Interestingly, a subpopulation of primary-like units, primary-like with a notch (PN), have been implicated in the phase-

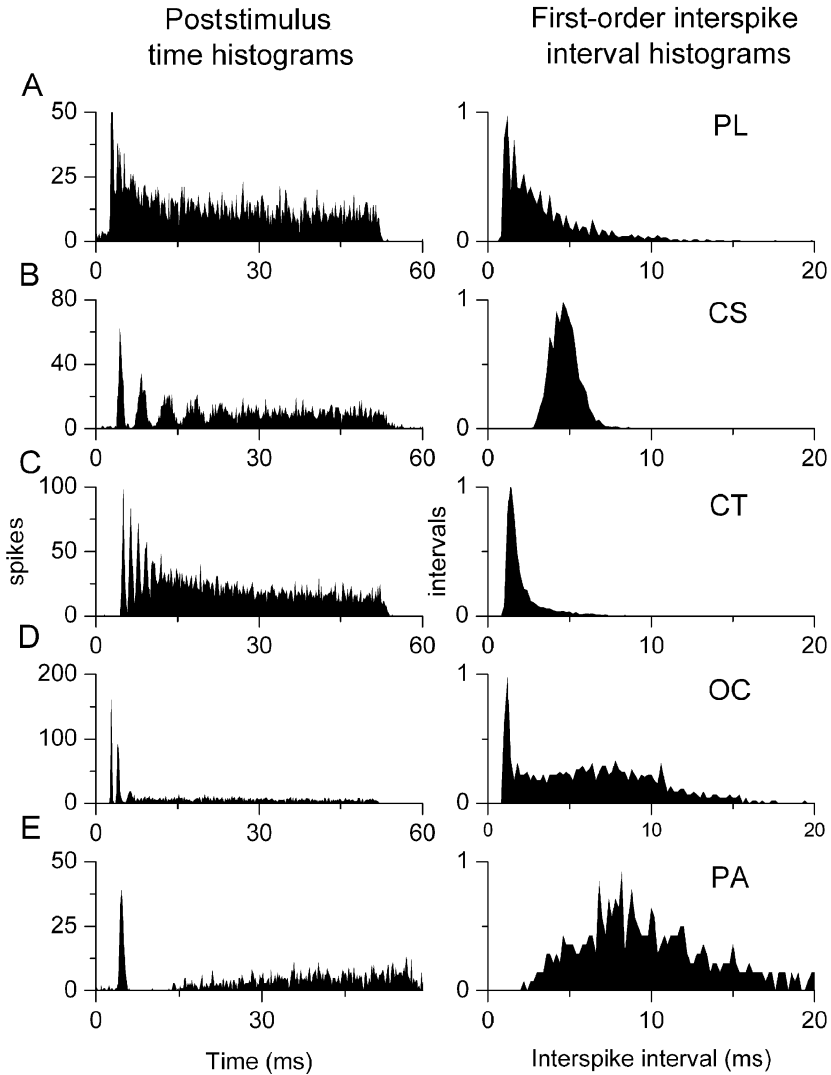


FIGURE 4.6. Temporal response properties of the main physiological response types in the mammalian cochlear nucleus. The poststimulus time histograms were obtained in response to 20 (A–C) or 50 dB (D and E) suprathreshold tone bursts at the unit's best frequency. The first-order interspike intervals were taken from the same spike trains used to generate the PSTHs. All recordings are from the cochlear nucleus of the anaesthetized guinea pig. CS = sustained chopper; CT = transient chopper; OC = onset chopper; PA = pauser; PL = primary-like.

opponency theory of level coding (Carney et al. 2002). The PN units are recorded from globular bushy cells in the ventral cochlear nucleus and may act as across-frequency coincidence detectors (Joris et al. 1994). This is consistent with the presence of inhibition, either lateral or centerband, in some primary-like and primary-like with notch units (Winter and Palmer 1990a; Caspary et al. 1994; Kopp-Scheinpflug et al. 2002).

## 2.2 *Chopper Units*

Chopper units are characterized by a regular chopping pattern in their PSTH (see Fig. 4.6B and C). There are at least two types, sustained (CS) and transient (CT), although some authors have subdivided them further (Blackburn and Sachs 1989). Sustained choppers are characterized by an extremely regular discharge pattern that does not change much on a presentation-to-presentation basis. This low variability has implicated them in the encoding of stimulus intensity (Shofner and Dye 1989), although interestingly such low variability does not appear to extend to their response to steady-state vowels where their variance is as great or greater than seen in their auditory nerve fiber input (May et al. 1998). The second type of chopper unit, CT, also shows a characteristic regularity in its discharge pattern but this regularity of response is maintained only over the first few milliseconds of response (i.e., it is transient). Most authors now readily group their recordings in the cochlear nucleus into these two groups but it is likely that this is convenient shorthand and the picture will get more complicated as we understand more about the recoding that goes on at this level.

In response to steady-state vowels it has been shown that both sustained and transient chopper units may represent the formant peaks in terms of their steady-state discharge rate at reasonably high sound levels (Blackburn and Sachs 1990). De facto one may extrapolate this result to the representation of the F0 in terms of mean spike discharge rate. However, it should be borne in mind that it is extremely difficult, if not impossible, to classify a unit as a chopper at low BFs (approximately  $< 0.4$  kHz) due to phase locking (i.e., despite the randomization of the stimulus starting phase you cannot tell whether the unit is phase locking or chopping). The good representation of formant peaks by chopper units at high sound levels has been interpreted in the light of the selective listening hypothesis (Winslow and Sachs 1988; Lai et al. 1994).

### 2.2.1 Selective Listening

A schematic diagram illustrating the principle of selective listening is shown in Figure 4.7. Auditory nerve fibers with high SRs form excitatory connections with the distal portion of the chopper unit dendrite whereas low/medium-SR fibers make excitatory contact on the more proximal parts of the dendrite or on the soma. These inputs share the same BF as the chopper unit. Off-BF inputs are hypothesized to make excitatory contacts with an inhibitory interneuron that then makes inhibitory connections with the chopper unit. The inhibitory inputs

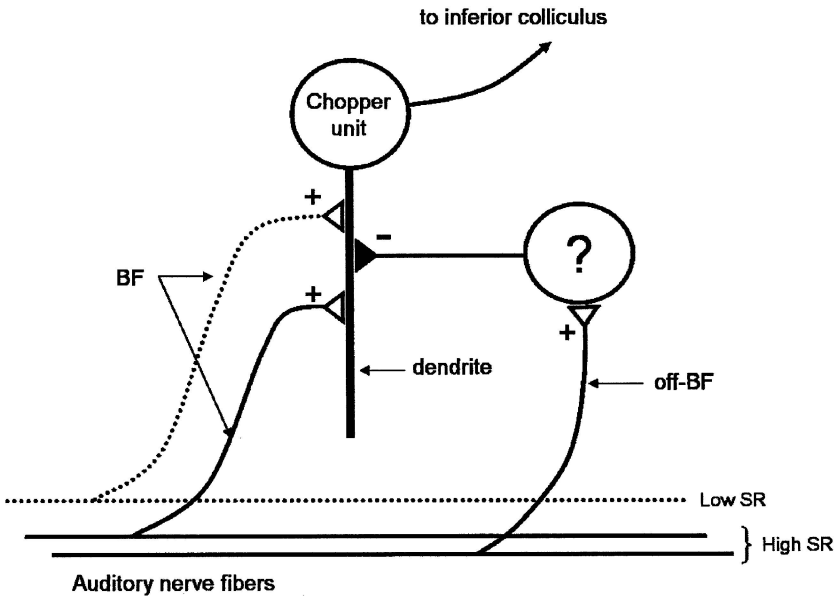


FIGURE 4.7. Schematic illustration of “selective listening” in the cochlear nucleus. It is hypothesized that chopper units in the cochlear nucleus can maintain a good representation of the harmonic structure of steady-state vowels at high sound levels by selectively listening to high-SR auditory nerve fibers at low sound levels and low-SR fibers at high sound levels. These units are able to do this with the assistance of an inhibitory interneuron (labeled “?” above) which receives off-BF input from high-SR fibers. This interneuron inhibits the response of on-BF, high-SR, auditory nerve fibers by effectively shunting their input on the dendrite enabling the proximally located low-SR auditory nerve fibers to dominate the response of the chopper unit.

are positioned between the high and low/medium-SR inputs and as such are on the direct path that current must take when flowing from the distal inputs to the soma. With this simple circuit one can see that at low stimulus levels the only active input to the chopper unit will come from the on-BF high-SR auditory nerve fibers. Increases in stimulus level will, through spread of excitation within the cochlea, activate the off-BF high-SR inputs, effectively eliminating any contribution from the more distally positioned on-BF fibers. At higher levels the contribution of the on-BF fibers will be ineffective while the input from the more proximally positioned low-SR fibers will be relatively unaffected. In this way the chopper unit may be thought of as selectively listening to high-SR fibers at low stimulus levels and low-SR fibers at high stimulus levels. Using a compartmental model of chopper units, Lai et al. (1994) were able to demonstrate the feasibility of such a circuit in reproducing the poststimulus time histograms and rate-level functions from “real” chopper units.

### 2.2.2 Representation of Concurrent Vowels

While the chopper units may be particularly well suited to represent the spectrum (and presumably F0) of complex sounds in terms of a mean rate code that is relatively insensitive to changes in sound level, it is also possible that they may play a role in the encoding of pitch in their temporal response to vowel sounds. As discussed in Section 1.3, differences in F0 between vowels aid in their separation and identification (Assmann and Summerfield 1990). Models of vowel separation generally involve two stages: (1) an estimation of the F0 of one or both of the vowels and (2) the selection or cancellation of the harmonics of one of the F0s. Chopper units provide a relatively poor representation of the formant peaks in terms of their temporal discharge patterns as they show greatly diminished phase locking in response to steady-state tones (Blackburn and Sachs 1989; Winter and Palmer 1990b). However, independent of unit BF, choppers show considerable modulation of their spike discharge pattern at the F0 of the vowel.

Keilson et al. (1997) have demonstrated that chopper units effectively provide a periodicity-tagged spectral representation, where mean discharge rate represents for stimulus energy and the temporal response represents the F0 of the stimulus that provides that energy. In the case of double vowels the mean rate output of the chopper unit is dominated by the profile of stimulus energy near the unit's BF while its temporal rate output is dominated by the F0 of the vowel with the larger amount of energy within its receptive field. Keilson et al. further speculated on the nature of a neural mechanism necessary to extract the F0 at higher levels of the auditory pathway. They suggested that the F0 could be detected by neurons that selectively respond to a limited range of periodicities, that is, are tuned to periodicity. Assuming a range of such periodicity-tuned filters, the two vowels would excite different places along the periodicity axis and the F0 would have been converted from a temporal to place representation. Just such a map has been hypothesized to exist in the central nucleus of the inferior colliculus (Langner and Schreiner 1988) and also the auditory cortex (Schulze and Langner 1999).

### 2.2.3 Representation of F0 by First-Order Interspike Intervals

Chopper units have also been hypothesized as the first stage of temporal processing which renders the need for an autocorrelation, that is, an all-order interspike interval analysis, unnecessary (Winter et al. 2001; Wiegube and Meddis 2004). For periodic sounds, chopping produces a decrease in the number of intervals that do not correspond to the chopping period and an increase in the number of intervals equal to the chopping period. If this chopping period equals the stimulus periodicity, the chopper-unit output is strongly locked to the stimulus period even when the stimulus period is not reflected in pronounced periodic envelope oscillations of the stimulus. In an array of chopper units with the same BF but with a range of chopping periods as hypothesized by Frisina et al. (1990)

and Kim et al. (1990b), the stimulus periodicity would be represented in an interval place code. Hewitt and Meddis (1994) demonstrated in a computer model of amplitude-modulation sensitivity of single units in the inferior colliculus how such a first-order interval-place code can be converted into a rate-place code in coincidence detector units presumably located in the central nucleus of the inferior colliculus. Winter et al. (2001) suggested that the model of Hewitt and Meddis (1994) for the coding of amplitude modulated pure tones could be extended to the case of more complex periodic stimuli where the periodicity is not so apparent in the stimulus envelope. In this suggested circuitry, the ventral cochlear nucleus (VCN) CS units could not only provide all-order to first-order conversion but also make temporal periodicity coding level insensitive. This hypothesis has recently been tested by Wiegand and Meddis (2004), who have shown that an array of CS units can indeed represent the pitch of complex sounds. Furthermore, owing to the relatively high sensitivity and steeply rising input-output functions of some CS units, with a dynamic range of about 20 dB (Rhode and Smith 1986; Blackburn and Sachs 1989; Winter and Palmer 1990a), their rate response saturates at relatively low levels. Above this saturation level the first-order interval code is level independent.

It should be recognized that this hypothesis requires a range of periodicity tuned units, encompassing the range of pitches, in each frequency channel. To date two studies (Frisina et al. 1990a,b; Kim et al. 1990b) have shown a distribution of best periodicity as a function of BF (see Fig. 4.8A and B). Kim and colleagues found best periodicities ranging from 100 to 500Hz in a population of units in the posteroventral and dorsal cochlear nucleus. No single unit type was found to encompass the whole range of best periodicities corresponding to the range of pitches perceived. A strong correlation between intrinsic oscillations and best periodicity was found in CS units. In contrast, Frisina et al. (1990) found little correlation between intrinsic chopping frequency and best periodicity (their best modulation frequency). However, Frisina et al. (1990) did not distinguish between CT and CS units and Kim et al. (1990) attributed the lack of correlation to the method of estimating a unit's intrinsic oscillation and/or a difference in unit type.

If the intrinsic oscillation seen in chopper units conveys the pitch of complex sounds then one must ask what is the range of pitches evoked by complex sounds. Recent studies suggest that the lower limit of pitch in humans is approximately 30 Hz (Krumbholz et al. 2000; Pressnitzer and Patterson 2001). This value is considerably lower than the lowest best oscillation frequencies seen in chopper units of the cat and guinea pig and this remains a problem for this theory of pitch encoding.

### 2.3 Onset Units

Onset units, as their name implies fire very precisely at the stimulus onset. They are now commonly classified into three groups; onset-I (OI), onset-chopper (OC—Fig. 4.6D) and onset-later activity (OL) (Rhode and Smith 1986; Winter

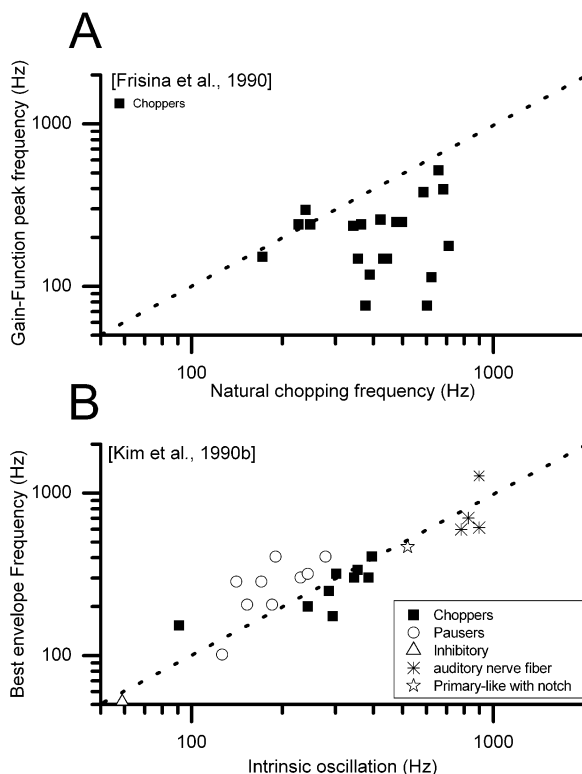


FIGURE 4.8. (A) Natural chopping frequency versus gain-function peak for single units in the gerbil cochlear nucleus. The natural chopping frequency was obtained by finding the time interval between the first four peaks of the response (Frisina et al. 1990). Note the apparent lack of a relationship between the two variables. (B) This is in contrast to the study of Kim et al. (1990b), who showed a close correspondence between intrinsic oscillation and best envelope frequency for single units in the DCN and PVCN of the cat. Note that chopper group consisted of 5 CS units and 2 CT units. The range of intrinsic oscillation in this study varied from 90 Hz to 400 Hz for the chopper group (Kim et al. 1990b). The *dotted line* in both plots is the line of unity.

and Palmer 1995). OC and OL units have a wide dynamic range. Assuming a first-order ISI code for pitch in the cochlear nucleus, OC and OL units, like CS units, may provide a conversion of higher-order to first-order intervals but the wide dynamic range of OC and OL units makes estimates of F0 from their responses level dependent. Whereas it is believed that CS units project to the inferior colliculus (Adams 1979; Smith et al. 1993), projection sites of OC units are still unclear and it may be possible that they act as interneurons in the CN (Joris and Smith 1998; Arnott et al. 2004). OC units represent the pitch of voiced speech sounds with remarkable fidelity and may respond to the ambig-



uous pitches of in-harmonic complexes in terms of inter-spike intervals (Kim et al. 1986; Palmer and Winter 1992, 1993; Rhode 1994, 1995). They also respond to the pitch of amplitude-modulated noise (see Plack and Oxenham, Chapter 2) in a manner that is similar to their response to 200% amplitude modulated tones, that is, three equal amplitude sinusoids using an interspike interval code (Rhode 1994). To explain the remarkable precision of spike timing in these units several authors have speculated that a form of across-frequency coincidence detection must be employed (Kim et al. 1986; Rhode and Smith, 1986; Kim et al. 1986; Palmer and Winter 1996).

If OC units are involved in encoding the pitch of complex sounds then it is important to know the termination site of their axonal projections. A few studies have examined this question, either directly or indirectly, and increasingly it appears that units with an OC PSTH shape may provide wideband inhibition both within the cochlear nuclear complex and also to the contralateral cochlear nucleus (Schofield 1995; Doucet and Ryugo 1997; Joris and Smith 1998; Arnott et al. 2004). Single units with an OC temporal adaptation pattern have been recorded from large multipolar cells within the ventral division of the cochlear nucleus (Smith and Rhode 1989). In the study of Smith and Rhode (1989) an axon of an OC unit was seen coursing through the DCN seemingly en-route to the exit pathways of the PVCN and DCN, the intermediate and dorsal acoustic striae. In slices of the mouse cochlear nucleus, Oertel et al. (1990) have identified a cell, which may be homologous to OC units in-vivo (their stellate-D cell). This cell had a dorsally projecting axon and formed connections with other multipolar cells in the ventral cochlear nucleus and with fusiform cells in the dorsal cochlear nucleus. In a study on the rat (*Rattus norvegicus*) cochlear nucleus, Doucet and Ryugo (1997) have shown that intracellular labeling in the fusiform layer of the dorsal cochlear nucleus labels large multipolar cells in the posteroventral cochlear nucleus; it is hypothesized that these large multipolar cells correspond to OC units. The connection of these cells in the cochlear nucleus is likely to be inhibitory as their terminals stain positively for glycine and are characterized by pleomorphic vesicles in their synaptic endings (Smith and Rhode 1989). In further studies, Doucet et al. (1999) have shown that large multipolar cells located in the PVCN project to the contralateral cochlear nucleus. The current anatomical information about OC units would seem to argue against them playing a role in the encoding of the pitch of complex sounds and suggests that other unit types (e.g., primary-like and choppers) may play a more pivotal role.

The OI unit type, firing mainly at stimulus onset, has also been implicated in the coding of periodicity (Godfrey et al. 1975, Rhode and Smith 1986; Oertel et al. 2000). There is increasing circumstantial evidence that the OI unit type corresponds to the octopus cell type. The dendrites of the octopus cell lie across the bundle of incoming auditory nerve fibers and they are thus ideally situated to receive information from a wide range of frequencies. Their morphological properties are reflected in their physiological responses to pure tones; they are very widely tuned and show strong temporal precision (Godfrey et al. 1975;

Rhode and Smith 1986; Winter and Palmer 1995). Recordings from octopus cells *in vitro* have demonstrated that, in responses to electrical shocks of the auditory nerve root, the synaptic potentials are very brief and their peaks are consistent within fractions of a millisecond. The firing rate of OI units can reach very high rates for low frequency tones (approximately 800 spikes/s); this is in comparison with the maximum discharge rate of auditory nerve fibers between 300 and 400 spikes/s. Octopus cells project to the contralateral ventral nucleus of the lateral lemniscus where they terminate with end-bulbs of Held (Adams 1997; Schofield and Cant 1997; Vater et al. 1997). The nuclei of the lateral lemniscus are located among the fiber tracts of the lateral lemniscus, a fiber tract that terminates in the inferior colliculus. Here, it is believed they synapse onto glycinergic cells which then project to the inferior colliculus. They are in a position to provide precisely timed inhibitory input to the inferior colliculus. While they respond with remarkable temporal precision (*in vitro*) and to high-frequency click trains several observations suggest they may not be well suited to encoding the pitch of complex sounds. For instance, Evans and Zhao (1998) have shown that units identified as OI did not respond well to random-phase harmonic complexes (RPH) but did respond well to cosine-phase harmonic complexes (CPH). However, these onset units were characterized by high BFs and it is possible that OI units with a low BF may be phase insensitive. This phase sensitivity has also been demonstrated in a couple of onset units in the chinchilla (*Chinchilla laniger*) AVCN by Shofner (1999), although the anatomical location of these units would appear to rule them out as coming from octopus cells.

## 2.4 The Representation of Periodicity in the DCN

So far we have concentrated on the responses of single units in the ventral division of the cochlear nucleus. However, the dorsal division may also be important for the temporal encoding of the pitch of complex sounds. Kim et al. (1990b) have shown that units classified as pause/build (Fig. 4.7E) show oscillatory behavior in response to single tones and amplitude-modulated stimuli. This appears to be similar to the oscillatory behavior of OC units. It has been hypothesized that units classified as OC may provide an inhibitory input to pause/build units (Nelken and Young 1994, Winter and Palmer 1995) and it is possible that the inhibitory input, if it is tuned, may impose an intrinsic rhythm upon the pause/build units. Of course, this does not preclude the possibility that both of these cell types generate their periodicity encoding *de novo*. Pause/build units also show a robust representation of amplitude modulation (AM) in the presence of background noise (Frisina et al. 1994). At levels of 0 dB S/N, the responses to the AM signal were as strong as those in quiet. Langner (1981, 1988) has also hypothesized that pause/build units are an important component in a model of periodicity analysis (see de Cheveigné, Chapter 6). Finally it is worth noting that a type IV unit in the DCN has been shown to respond to either AM or quasi-frequency-modulated stimuli (QFM) in an almost identical

manner indicating an insensitivity to component phase (Rhode 1994). Although units in the auditory nerve may be sensitive to component phase some units in the cochlear nucleus seem relatively impervious to alterations in component phase.

## *2.5 Responses of Single Units in the Cochlear Nucleus to Iterated Rippled Noise*

Iterated rippled noise was first introduced into auditory psychophysics by Bilsen and Ritsma (1969/70) following the discovery of a description of a stimulus very much like iterated rippled noise by Christian Huygens in the seventeenth century. This stimulus is further described by both Plack and Oxenham (Chapter 2), Shofner (Chapter 3), and Griffiths (Chapter 5) and is described only in brief here.

Rippled noise (RN) can be produced from white noise by delaying a copy of the noise by  $d$  ms and adding the delayed noise back to the original. Iterated rippled noise (IRN) is produced by repeating the delay and add process  $n$  times, and it is referred to as  $\text{IRN}(d,n)$ . The delay-and-add process introduces temporal regularity into the fine-structure of the noise (Figures 4.9A and B) which is revealed by peaks in the autocorrelation function of the wave (Figures 4.9E and F). It also introduces a “ripple” into the long-term power spectrum of the wave (Figures 4.9C and D). Note, however, that the resolution of the spectral analysis performed in the cochlea is proportional to frequency and so high-frequency peaks merge in the internal tonotopic representation. Simulations of the processing of IRN (Griffiths et al. 1998) do not show resolved peaks above about the sixth harmonic (this obviously crucially depends on the frequency resolution of the modeled filter bank). If one uses either Shamma’s lateral inhibitory network (Shamma 1985a,b) or a filterbank based on the bandwidths estimated by Shera et al. (2002) then a greater number of resolved harmonics will be present. It has been argued (Yost et al. 1996) that the pitch of IRN is best represented by the position of the first peak in the autocorrelation of the waveform.

The iteration process does produce some amplitude modulation (AM) in individual frequency channels; however, the modulation has a different phase in each presentation, it has different phases in different channels, and the phase drifts continuously in every channel. As a result, the stimulus does not have the pronounced envelope modulation typical of traditional pitch producing stimuli such as AM tones or AM noise. The form of the modulation is important for two reasons: First, it means that the stimulus precludes a simple phase-locked analysis of the spike discharges as the phase of the modulation is changing over time and between channels. Secondly, it reduces the chances that the results will be confounded by aural distortion products. When AM is applied to a tone or band of noise, even-order distortion in the cochlea generates a relatively strong distortion component on the basilar membrane at the modulation frequency (Wiegand and Patterson 1999). As a result, when the response of a unit exhibits activity at the modulation frequency, it is difficult to determine whether

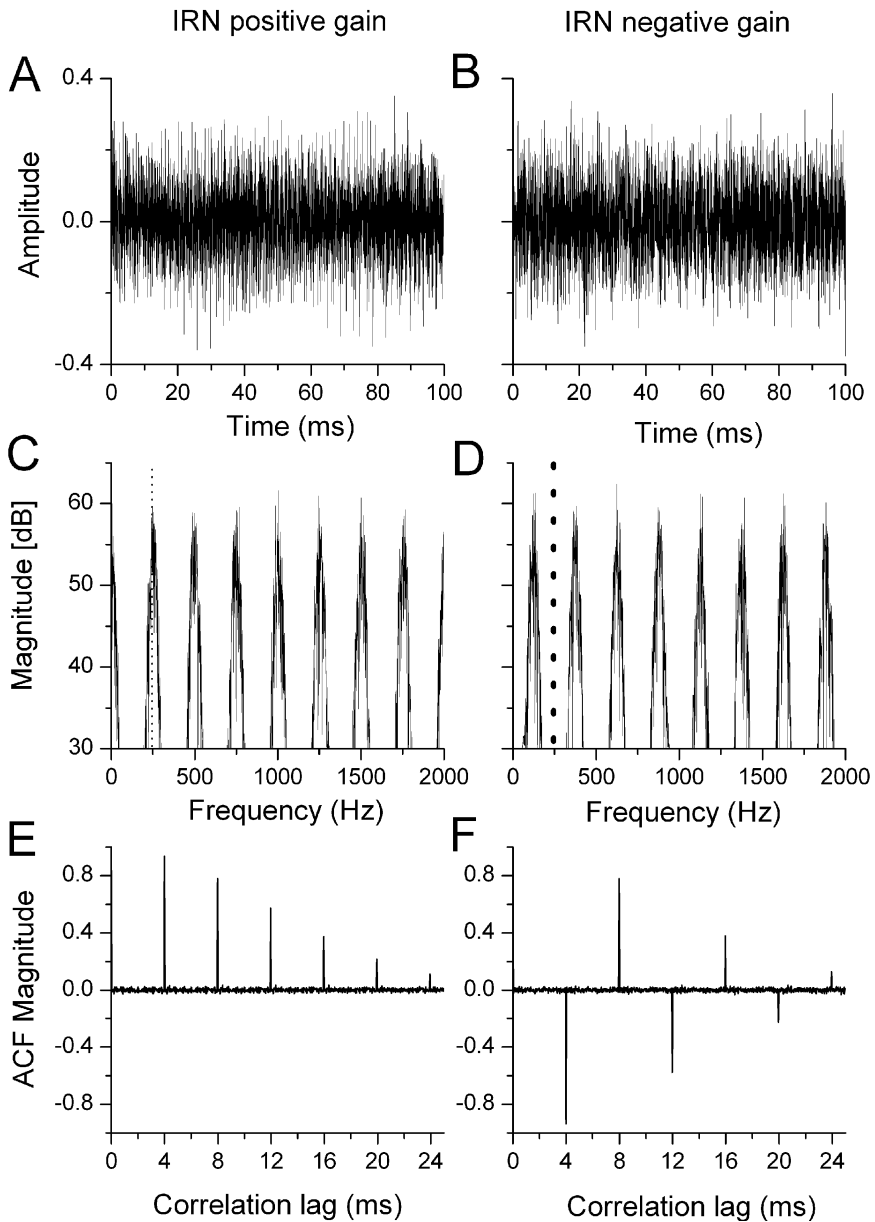


FIGURE 4.9. Iterated rippled noise calculated with positive (*left column*) and negative (*right column*) gain. **A** and **B** are the waveforms for 16 iterations and a delay of 4 ms. The spectra are shown in panels **C** and **D**. Note the lack of a distinct peak at the reciprocal of the delay in the spectrum for the negative gain stimulus. In **E** and **F** the autocorrelation of the waveform illustrates that the largest positive peak is at the delay for the positive gain stimulus whereas it is at twice the delay for the negative gain stimulus.

the response represents central extraction of the modulation information from high-frequency channels, or a direct response to a distortion component at the modulation frequency on the basilar membrane.

In a physiological study Shofner (1991) investigated the temporal representation of *rippled* noise (RN) in the anteroventral cochlear nucleus of the chinchilla. He concluded that, while primary-like units seemed to preserve the RN fine structure, chopper units code the quasi-periodicity only in the stimulus envelope. This was based on the finding that rippled noise which was delayed and added (the gain,  $g$ , in the delay-and-add loop equals 1) was coded in the same way as rippled noise which was delayed and subtracted ( $g$  equals  $-1$ ). Shofner (1999) investigated the temporal response characteristics of the chinchilla cochlear nucleus units in response to infinitely iterated rippled noise (IIRN) with a  $g$  of 0.89 or  $-0.89$  in the delay-and-add loop. As for the rippled-noise results, Shofner (1999) concluded that while PL units do preserve the difference between IIRN with positive and negative  $g$ , this was not the case for chopper units. IIRN with positive and negative “ $g$ ” share the same envelope features but differ in their temporal fine structure. This argument would be consistent with the idea that chopper units are envelope responders and PL units are driven by fine-structure information and is illustrated in Figure 4.10, which shows the responses of a primary-like and chopper unit to IRN(–) and IRN(+) recorded from the guinea pig cochlear nucleus. However, Shofner (1999) showed responses of two PL units with BFs of 0.85 and 4.63 kHz. Whereas the autocorrelation of the low-BF response reflected the stimulus autocorrelation, the high-BF units’ autocorrelation was the same irrespective of the sign of the IIRN gain as it was the case for the 2.43-kHz chopper unit shown in Shofner (1999). Thus, the temporal representation of IIRN may be more dominated by BF than unit type. This interpretation would also be more in line with the perception of the stimuli. For a fixed delay of 4ms, the pitch difference between positive and negative  $g$  is an octave only when the low harmonics are presented. When the stimuli are high-pass filtered, the pitch difference is much smaller and more on the order of 10%, as it is observed for rippled noise (Bilsen and Ritsma 1969/70). It is therefore possible that chopper units with low BFs may be capable of preserving differences related to the sign of  $g$  in their temporal response properties as far as these are established perceptually. Preliminary evidence supporting this idea has been found in the cochlear nucleus of the guinea pig for transient chopper units with BFs below 1.1 kHz (Verhey et al. 2004).

### 3. The Inferior Colliculus

If the cochlear nucleus is an obligatory synapse for the auditory nerve then equally the inferior colliculus (IC) may be considered an obligatory synapse for the overwhelming majority of cells from the nuclei below it in the auditory brainstem. All nuclei except the contralateral ventral nucleus of the lateral lemniscus send projections to the central nucleus of the IC (ICC) on both sides.

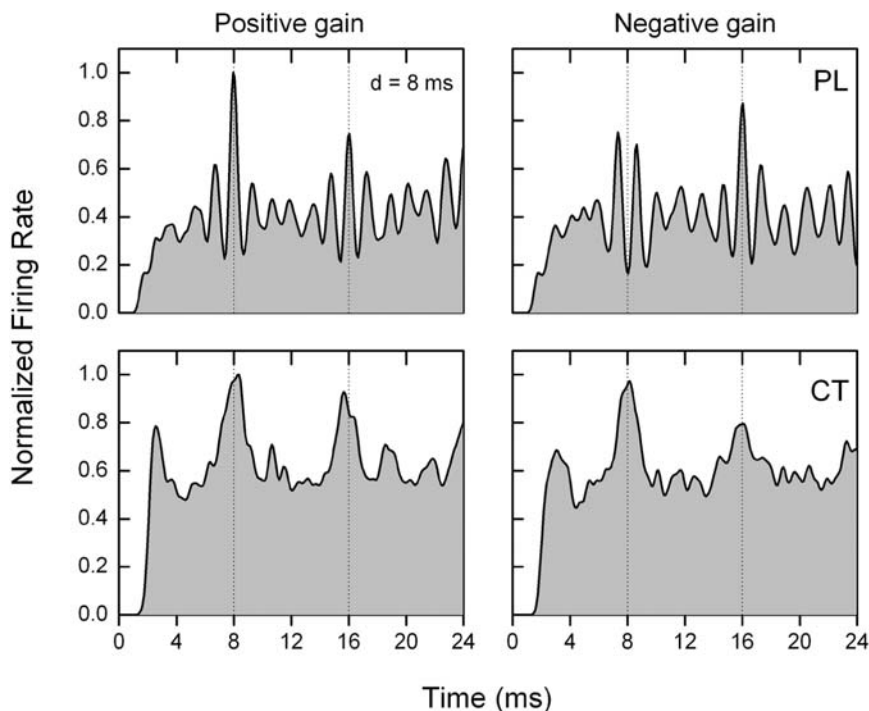


FIGURE 4.10. Neural autocorrelation functions in response to iterated rippled noise with a delay ( $d$ ) of 8 ms and a positive (*left column*) or negative gain (*right column*). For the primary-like unit (upper row—BF = 0.84 kHz) the largest peak is found at a  $d$  = 8 ms for the IRN (+) while for the IRN (−) condition the largest peak is found at  $d$  = 16 ms. This is consistent with the perception of these two stimuli. In contrast, the neural autocorrelations for the transient chopper unit (lower row—BF = 3.6 kHz) are almost identical, with the largest peak occurring at 8 ms in each case. Both units were recorded from the ventral cochlear nucleus of the anesthetized guinea pig.

Most axons in the lateral lemniscus synapse in the ICC with relatively few bypassing this nucleus and terminating in the thalamus. The IC is composed of several subdivisions that can be distinguished by cytoarchitecture (Rockel and Jones 1973a,b; Willard and Ryugo 1983; Oliver and Morest 1984). The ICC contains two main cell types; principal cells, which are bitufted fusiform or disk-shaped cells, make up more than 70% and their dendritic trees are oriented with their long axis parallel to the ascending lemniscal axons. The thickness of the dendritic tree determines the width of the lamina (70 to 150  $\mu\text{m}$ ). Multipolar or stellate cells of various kinds have irregular dendritic trees or those that are oriented mainly orthogonal to those of the principal cells and lemniscal axons. Like the cochlear nucleus, the ICC is organized tonotopically; low frequencies are located dorsally while high frequencies are found more ventrally

(Merzenich et al. 1975). However, the responses to single tones are considerably more complex, with 60% of the neurons responding to stimuli in either ear. Despite being the most accessible nucleus in the auditory brainstem, surprisingly little has been studied about its representation of pitch. In contrast, information has been gathered on its ability to represent sinusoidal amplitude modulation (see Joris et al. 2004 for a review) and it is to these data that we turn for an indication about how this area of the pathway may respond to pitch (see Section 3.1). In contrast to the responses of single fibers in the auditory nerve, many units in the IC are characterized by nonmonotonic rate-level functions (Semple and Kitzes 1987; Ehret and Merzenich 1988; Rees and Palmer 1988; Irvine and Gago 1990). There appears to be a continuous distribution of rate-level function shapes from monotonic to highly nonmonotonic (Irvine and Gago 1990) and consequently the number of units classified as either monotonic or nonmonotonic depends on the criterion chosen. The nonmonotonocities have implications for the type of units that may be involved in coding sound level. For instance, Ehret and Merzenich averaged the discharge rates of a population of units from the ICC and found that there was essentially no change in discharge rate output over a wide range of stimulus levels; however, it is possible that more central nuclei use only those units that are monotonic in estimating sound level.

Alternatively, it has been argued that sound level is represented by a series of neurons with “best-SPLs,” that is, they are sharply nonmonotonic (Brugge and Merzenich 1973; Phillips and Orman 1984). Therefore a place code would exist for sound level with each particular place responding only to a certain SPL. However, doubt has been cast on this idea by Ehret and Merzenich, who have shown that the “best-SPL” is dependent on the spectral content of the stimulus, that is, they peaked at different levels for tones and noise. It is clear from the foregoing studies that the encoding of the frequency of a pure tone at high sound levels is not a simple affair. It is often argued that sound level is coded by neurons with different thresholds but the evidence for this is, at best, sparse and further discoveries await before we can be confident how stimulus level is encoded at this level of the auditory pathway.

### *3.1 Periodicity Tuning*

Many neurons in the central nucleus of the IC show a bandpass selectivity to amplitude modulation, either in their mean discharge rate or phase-locked output (Langner and Schreiner 1988; Schreiner and Langner 1988; Rees and Palmer 1989; Krishna and Semple 2000; Langner et al. 2002). The periodicity information is present up to several hundred Hertz in a temporal code but is present up to frequencies of 1000 Hz in a mean rate discharge code (Fig. 4.11). In many neurons with a bandpass modulation transfer function (MTF) a burst-like intrinsic oscillation is triggered at signal onset and often at each modulation cycle. In contrast to CS units in the cochlear nucleus these intrinsic oscillations do not equal the unit’s best modulation frequency (BMF) and this presents dif-

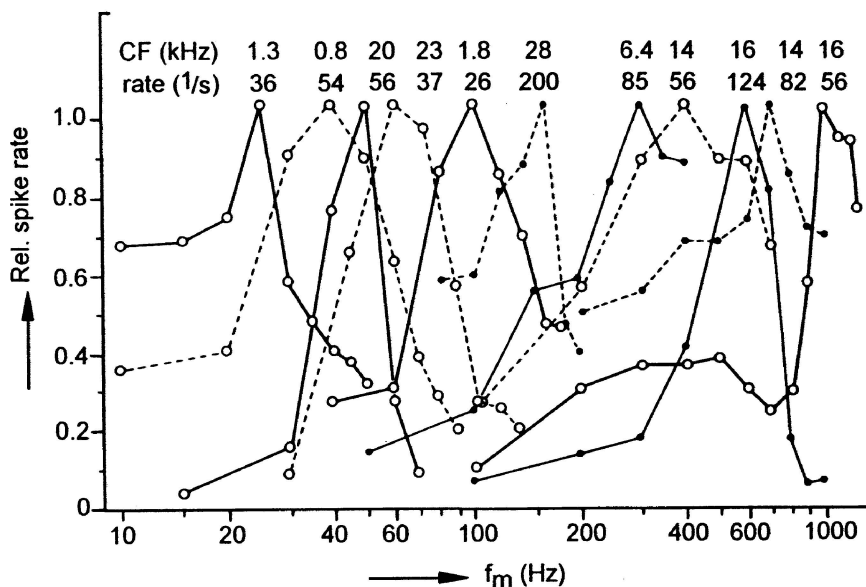


FIGURE 4.11. Arguably the most famous result from neural recordings in the central nucleus of the IC. Each curve represents a modulation transfer function for amplitude modulated tones. The range of BFs are given at the top of each curve along with the maximum output of each unit. Note the range of best modulation frequencies extends from 20 Hz to 1000 Hz. This result was obtained in the cat (Langner and Schreiner 1988) although a similar result has also been reported in the gerbil (Langner et al. 2002).

faculties for the model by Hewitt and Meddis (1994), who proposed that sustained chopper units contact IC units and, through coincidence detection, imposed their BMF on units in the ICC.

Langner and colleagues (Hose et al. 1987; Langner and Schreiner 1988; Langner et al. 2002) have argued that there is a map of BMF that runs orthogonal to the pure tone frequency map, however, many criticisms are often levied at this map, including: (1) MTFs are too broad to support the fine pitch discriminations that we can make psychophysically; (2) the MTFs become broadband at higher sound levels even though our perception of the pitch of complex sounds changes very little; and (3) the range of BMFs is not sufficient to support the encoding of pitch much above 1200 Hz. In response to these criticisms I know of no quantitative model that has tried to use these broad filters to explain data on pitch discrimination but our discrimination of color is possible with the use of just three, broadly tuned filters. The use of broadly tuned filters has also recently been proposed as a means for encoding interaural time differences in mammals (see McAlpine and Grothe 2003 for a review) and therefore the use of relatively broad filters could be a common feature in neural systems. While



the data indicate that the shape of the MTFs is level dependent there is, nevertheless, a wide variation in threshold of single units in the IC and it is possible that, similar to the auditory nerve, one group of units is used at one level and another group at higher levels. Finally, in response to point (3) above this issue was addressed by Langner et al. (2002), who argued that the reason for not finding BMFs greater than 1200 Hz was largely a sampling issue.

In a study looking at the ability of single units in the IC to integrate periodicity information Biebel and Langner (2002) showed that neurons could respond to modulation even when the carrier frequency was positioned far from the excitatory part of the unit's receptive field. However, one must be cautious in interpreting these results because of the possibility of distortion. McAlpine (2004) has demonstrated that some neurons in the IC do indeed respond to the distortion produced by high-pass-filtered complex stimuli. Notwithstanding the criticisms faced by Langner's model of periodicity coding it would be interesting to test this model with more complex stimuli. What happens to the periodicity maps when using stimuli other than AM tones? For instance, how do neurons in the ICC respond to iterated rippled noise, a stimulus with a distinct pitch but a greatly reduced modulation? Although neurons in the IC respond to the missing fundamental is this simply a response to distortion? A thorough, systematic study is now required to look at the responses of IC neurons to a variety of pitch producing stimuli along the lines of those used by Carianni and Delgutte (1996a,b) in the auditory nerve. The IC is also an obvious place to look for physiological correlates of binaural pitches. Are the cells involved in binaural pitch the same ones involved in monaural pitch perception or are monaural and binaural pitches compared at some more central (cortical?) area.

## 4. The Auditory Cortex

The role of the auditory cortex in auditory perception can, perhaps, best be described as enigmatic. Lesion evidence (Whitfield 1980) and, more recently, brain-imaging studies (see Griffiths, Chapter 5) have implicated the auditory cortex in the representation of the pitch of complex sounds but corresponding evidence from single unit studies has not been easy to demonstrate.

While great progress has been made in our understanding of the visual cortices, corresponding progress in the auditory cortices has been, at best, slow. It is not unreasonable to hypothesize the presence of neurons in the auditory cortex that respond to the pitch of both simple and complex sounds. Two questions must be answered: (1) Where in the auditory cortex does this take place? and (2) (perhaps more importantly), When is a "pitch" neuron not a "pitch" neuron, that is, what properties would we expect a pitch neuron to have? In response to the first question brain imaging studies have now given us a guide as to where to look. In response to the second question one could stipulate that in order for a neuron to be classified as a "pitch neuron" it would have to respond to the pitch of the stimulus irrespective of its spectral content, be relatively level in-

dependent, be duration sensitive, and with responses that are highly correlated with the perceived F0 as observed behaviorally. This section is confined to those studies that have looked at the representation of pitch measuring the direct electrical activity of single and multi-neurons. For a discussion of the numerous pieces of work using imaging techniques such as fMRI and MEG the reader is referred to the chapter by Griffiths (Chapter 5).

#### *4.1 The Representation of the Frequency and Level of Single Tones*

In anesthetized animals most units in the auditory cortex respond to the onset of a stimulus. Cortical neurons also fire spontaneously, although the spontaneous discharge rate is very low in anesthetized animals. While many units in the auditory cortex are characterized by “V”-shaped excitatory receptive fields near threshold, at levels well above BF threshold the receptive fields can vary considerably. Indeed many units have circumscribed areas of excitation and therefore could be said to be characterized by a sharp filter for both frequency and intensity (Phillips et al. 1985). Merzenich et al. (1975) demonstrated that precise tonotopic maps could be measured in the primary auditory cortex along the anterior–posterior axis. This tonotopicity was strongest in the primary auditory cortex. However, this description is now known to be far more complicated; for instance, Phillips et al. (1994) have shown that for low-level tones most responses from AI neurons occurred along the appropriate single isofrequency contour but that at high sound levels neurons outside the isofrequency respond to the single tones and other areas along the same isofrequency contour cease to respond (presumably because of inhibition). The coding of stimulus level is just as complicated although Heil et al. (1994) have shown that it is possible to combine single-unit information from isofrequency contours in A1 of the cat to offer an explanation of intensity discrimination and the encoding of loudness in humans.

Whether a neuron responds to a sound also depends on its novelty. Ulanovsky et al. (2003) have shown that neurons in the primary auditory cortex of cat responded more strongly to rarely presented sounds than to more commonly presented sounds. Of relevance to the representation of the pitch of simple sounds is that the frequency resolution for rare sounds was an order of magnitude better than receptive field bandwidths in primary auditory cortex. Importantly, they could demonstrate that such hyperacuity was not present at the level of the thalamus.

Increasing evidence suggests that the auditory cortex does not contain a static representation of sound but is undergoing constant reorganization in the face of changing inputs. It is now well established that the tonotopic maps seen in response to low-level tones can be altered either by cochlear lesions (Roberston and Irvine 1989), pairing sound frequencies with a reward (Edeline 1998) or with electrical stimulation of the basal forebrain (Kilgard and Merzenich 1998). A cell’s receptive field can also be altered during the performance of an auditory

discrimination task (Fritz et al. 2003). The changes, usually an increase in excitation or decrease in inhibition around the frequency of interest, occurred over minutes and did not require electrical stimulation or other physiological/pharmacological insults. This suggests that the cortex is able to reorganize very rapidly enabling a larger population of neurons to respond to the sound of interest. Of course many questions remain: When trying to detect a 1-kHz tone, cortical neurons that were previously tuned to neighboring frequencies now become tuned to the 1-kHz tone so does this make it harder to hear a nearby frequency? Or do the 30% of neurons that do not change their responses still maintain an orderly tonotopic map?

While the responses of the auditory cortex depend strongly on such things as level, electrical stimulation, alteration of inputs, and even novelty, a consensus is, nevertheless, emerging about the general organization of the auditory cortical areas; a tonotopic core is surrounded by a belt of cortex that is less tonotopically organized, which in turn is surrounded by cortex that is weakly tonotopic at best (see Semple and Scott 2003 for a review). The auditory cortical areas also correlate well with increasing stimulus complexity with the inner areas responding well to tonal stimuli while the outer areas respond best to more complex stimuli (Patterson et al. 2002; Wessinger et al. 2001).

## 4.2 The Representation of the F0 in Complex Sounds

A study in the awake macaque (*Macaca fascicularis*) has conspicuously failed to find a representation (Tomlinson and Schwartz 1990) of the missing F0. The responses of single units were determined by the relationship between the stimulus spectrum and the unit BF. This result is at odds with human imaging studies using MEG which have shown that there is a topographic representation of F0 in A1 (Pantev et al. 1989). While one cannot rule out an obvious species difference, the macaques were able to perceive the missing fundamental stimulus (Tomlinson and Schwartz 1988) and therefore it is possible that the representation of F0 is carried out by populations of units rather than single units in the A1. In an attempt to resolve this discrepancy, Fishman et al. (1998) examined the representation of the F0 of harmonic complexes missing the F0 by measuring either multi-unit activity (MUA) or current source density (CSD) analysis. In line with the single-unit data, Fishman et al. (1998) found that the responses in A1 were dominated by the spectral content of the stimulus rather than by its pitch.

In a related study, also using multi-unit analyses and current source density analysis, Steinschneider et al. (1998) have shown that the encoding of alternating polarity click trains in the macaque primary auditory cortex was dependent on click rate. Click rates between 100 and 200 Hz were represented by high-BF regions of A1 through phase-locked activity in the MUA and CSD and was independent of pulse polarity. In contrast, encoding of spectral features was found in low BF regions with resolution of both F0 and its harmonics being manifest by peaks of activity determined by the tonotopic organization of the

recording sites. Psychophysically, the pitch of click trains with pulse rates less than 100 Hz is determined by the pulse rate and is independent of pulse polarity. In contrast, the pitch of click trains with pulse rates greater than 200 Hz is determined by the F0 dependent on pulse polarity. The similarity between the psychophysics and physiology led Steinschneider et al. (1998) to conclude that the data supported the existence of two pitch mechanisms (e.g., Carlyon and Shackleton 1994); one using resolved harmonics and the other using unresolved harmonics. Two populations of neurons have been also found in the primary auditory cortex of the awake marmoset (*Callithrix jacchus jacchus*) in response to time-varying stimuli (Lu et al. 2001). One population responded to click trains with long interclick intervals (ICIs) with stimulus-locked discharges whereas a second population responded with nonstimulus locked discharges to click trains with short ICIs. Combined, the two populations were able to represent a range of ICIs from 3 to 100 ms. When plotted as a cumulative sum of the histograms of the distribution of synchronization boundaries (Fig. 4.12) there is a clear deflection point of the stimulus-locked (or synchronized) distribution near 25 ms. This is near to the lower limit of pitch at 30 ms.

### 4.3 Integration Beyond the Classical Receptive Field

The concept of the classical receptive field is interpreted with caution at the level of the cortex. In addition to the complex inhibitory inputs that can be measured from single units, it is increasingly apparent that the bandwidths of single units can be extremely broad, and tuning is not always obvious when measured using conventional techniques. Schulze and Langner (1999) have demonstrated that single units in the auditory cortex of the gerbil (*Meriones unguiculatus*) will respond to stimuli whose spectrum is completely outside of the single tone excitatory response field. They found that approximately 75% of units with BFs less than 3 kHz would respond to SAM tones with all components outside the excitatory receptive field. Again the problem of cochlear distortion generating the response to F0 needs to be addressed although preliminary data indicate that distortion is more of a problem at the level of the IC than the auditory cortex (Schrottge et al. 2004).

Using a combination of conventional microelectrode techniques and optical imaging, Schulze et al. (2002) have shown that best periodicity is represented in a circular (or horseshoe)-like fashion on the surface of the cortex. The best periodicities were mapped out using a single SAM tone burst with a fixed carrier frequency (8 kHz) and therefore the problem of distortion needs to be addressed. Schulze et al. (2002) speculate that such an anatomical arrangement could underlie the need for cellular interactions not found or plausible in the linear tonotopic map of frequency.

Several authors have reported the presence of multi-peaked receptive fields in the auditory cortex of the cat (e.g., Sutter and Schreiner 1991). This result has been replicated in the awake marmoset by Kadia and Wang (2003) who found approximately 20% of neurons with multi-peaked receptive fields. The excitatory

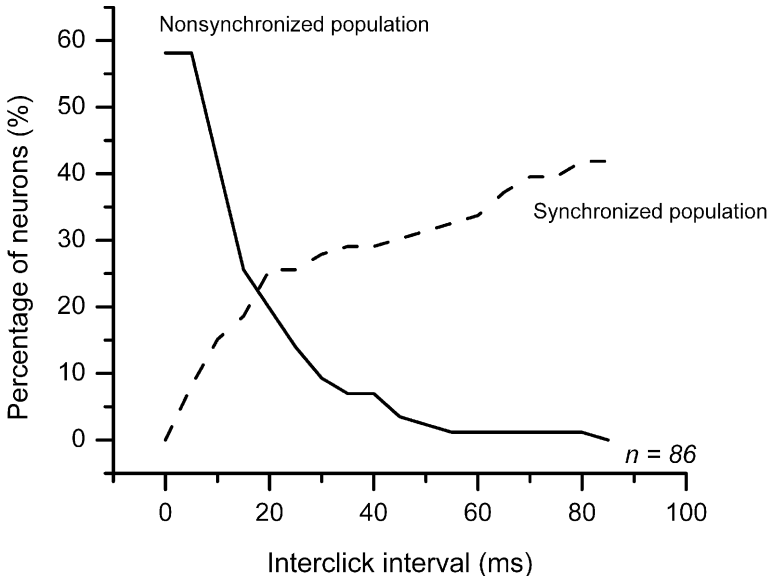


FIGURE 4.12. The representation of interclick interval in terms of temporal (synchronized) or rate codes (nonsynchronized). The *dashed line* shows the percentage of neurons with synchronization boundaries less than or equal to a given ICI. The solid line shows the percentage of neurons with rate responses greater than or equal to a given ICI. Note, however, that most neurons, synchronized and non-synchronized, preferred interclick intervals less than 20 ms. Data are replotted from Lu et al. (2001).

spectral peaks were often harmonically related and single units could show facilitation by combinations of tones selected to be in the positions of the excitatory peaks measured in the two-tone response areas. Unfortunately, the majority of multi-peaked units in both the cat and marmoset had BFs greater than 5 kHz—an obvious problem for all theories of pitch perception. However, the concurrent presentation of harmonically related frequencies gives a perception of a fused, single, harmonic complex tone and multi-peaked neurons could be a possible neural substrate subserving such perceptual observations.

#### 4.4 Descending Systems

The auditory cortex generally receives its auditory input from the medial geniculate body (MGB) located in the thalamus and more specifically to layers 3 and 4 (see Smith and Spirou, Chapter 2, in *Springer Handbook of Auditory Research*, Vol. 15). In contrast, layer 6 is a major source of descending input to the thalamus. In the visual system the number of synapses made on cells in the lateral geniculate nucleus (the visual homolog of the MGB) exceeds the number of synapses coming from the periphery (e.g., Erisir et al. 1997). These direct

connections are mainly excitatory in manner; inhibition is provided through the action of GABAergic interneurons within the thalamus. In the mustached bat (*Pteronotus parnellii parnellii*) activation of a localized area of cortex enhances the responsiveness of individual neurons in the MGB but only if the BFs of the cortical neurons matches the BFs of their thalamic targets. If the BFs do not match, the responsiveness of the thalamic neurons decreases. Furthermore, the receptive fields of mismatched MGB neurons shift away from the BF of the stimulated region of the cortex. Thus it appears that the cortex is able to “tune” its own input. This phenomenon has been termed egocentric selection (Suga et al. 2000). The implications for the encoding of the pitch of simple and complex sounds are unclear but it is possible that the cortex makes an initial assessment that the frequency or F0 of that area of cortex is present in the sensory signal. The cortex then amplifies the response of neurons in the thalamus that represent the predicted frequency or F0 while inhibiting the responses of thalamic neurons that do not, thus enabling the cortex to increase the signal-to-noise ratio of its own input. A similar result has been found in single units in the IC following electrical stimulation of the auditory cortex in the house mouse (*Mus domesticus*). Yan and Ehret (2002) have shown that the BFs of IC units may shift toward the BF of the stimulated cortical area. If the BFs of the stimulated cortical area and the units in the IC did not match, the thresholds of IC units were elevated and the dynamic range was reduced. Like the studies in the cortex, processing of sound in the center of cortical feedback can be enhanced while processing in the surround is suppressed.

## 5. Summary

It would be premature to claim that we knew how pitch is represented in the mammalian auditory pathway. Even at the level of the auditory nerve, several questions remain. For example, the relationship between SR, threshold, and dynamic range appears to hold over a variety of animals, but does the human auditory nerve have the same distribution of fiber types according to SR and threshold? How well do single fibers in the auditory nerve of humans phase lock? What is their corner frequency and cutoff slope? Many models use the decline of phase locking with frequency as measured in the cat; however, phase locking in humans may more closely resemble that found in the guinea pig, or even the barn owl! Given their high thresholds and relatively wide dynamic ranges, auditory nerve fibers with low SRs generated a lot of interest in their ability to represent F0 at high sound levels and, perhaps more importantly, in the presence of background noise. However, it is possible that fibers with low SRs may be more involved in cochlear feedback loops. This idea has received support from the observation that low-SR fibers terminate in the granule cell area of the cochlear nucleus (Liberman 1991, 1993) and also the similarity of the rate-level functions of olivocochlear efferent fibers and low-SR primary af-

ferent fibers (Liberman 1988). Until we are able to selectively eliminate the contribution of low-SR fibers to perception, their function will remain obscure. Finally, how sharply tuned are single auditory nerve fibers in humans? While we may be getting closer to an answer to this question (e.g., Shera et al. 2002; Oxenham and Shera 2003), until we can record the responses from the intact (and nondiseased) auditory nerve fibers in humans, the answers to these questions will probably remain elusive and the subject of constant speculation.

At present, neurophysiological evidence would appear to support an interspike interval representation of F0 at the level of the auditory nerve and cochlear nucleus (Evans 1978; Javel 1980; Rhode 1995; Cariani and Delgutte 1996a,b), although even this representation runs into trouble with the click trains from hell! At the level of the cochlear nucleus, under normal conditions, primary-like units are best able to preserve the temporal input from the auditory nerve and are thus good candidates to represent the temporal fine structure of the pitch of complex sounds. However, as judged by their anatomical projections, they are more likely to be involved in the encoding of space (although this does not preclude them from encoding both pitch and space). Chopper units in the cochlear nucleus have been proposed as a stage in the conversion from all-order ISIs to first order ISIs by acting as a series of resonators, each with their own preferred resonant frequency (Hewitt and Meddis 1994; Wiegrebe and Winter 2001; Wiegrebe and Meddis 2004). At the level of the cochlear nucleus it will also be important to test the competing hypotheses for how the level of a low frequency sound is encoded. Kim et al. (1991) have demonstrated that a population of chopper units is able to represent a low-frequency tone by a peak at the appropriate place in the rate-place profile. This peak was present at sound levels where most high-SR auditory nerve fibers had saturated and it was suggested that the chopper units were responding to the unsaturated low-SR inputs. This result is consistent with the selective listening hypothesis but are cells in the cochlear nucleus really able to selectively listen to low SR auditory nerve fibers or do they act as phase-opponent coincidence detectors? A particular attraction of the phase-opponency model is its ability to explain the paradoxically poor temporal sensitivity of patients with cochlear implants. Although auditory nerve fibers are well synchronized to electrical stimulation the phase delays normally associated with acoustic stimulation will be greatly altered, leading to disrupted spatiotemporal patterns of activity arriving in the cochlear nucleus.

Recent studies by May et al. (1998) have shown that a good representation of the formant peaks of steady-state vowels may be found in the discharges of primary-like and chopper units. Furthermore, the efferent system appears to help maintain a good mean-rate representation of complex sounds in background noise. However, many questions remain: Are the efferents equally effective at low frequencies—that is, the frequencies normally associated with pitch? Under what conditions is the olivocochlear system normally active?

Surprisingly, given their excellent response to the periodicity of many complex sounds, onset-chopper units are unlikely to be involved in pitch coding as

they project only within and between cochlear nuclei and are most likely inhibitory in action. We still do not know the precise projections of the different unit types in the cochlear nucleus. For instance, do the different types of chopper unit project to different targets in the IC? What cells do OC units contact in the contralateral cochlear nucleus? Are all the contralaterally projecting cells OC units? Can OC units project to higher levels in the auditory pathway? Is there a difference between OC and OL units? What role can OI units play in the encoding of pitch? It is clear from these questions that we still lack a complete understanding of the representation of pitch even at the level of the cochlear nucleus.

Information about F0 in the temporal discharge properties of single units probably disappears as one ascends the auditory pathway and it becomes necessary to search for a time to place conversion somewhere along the pathway. One such possibility is the modulation filter bank in the IC. Regrettably, this map has yet to be found by other groups. A related idea was suggested by Wiegrefe and Winter (2001), who adapted a previous observation about the encoding of AM (Kim et al. 1990b; Hewitt and Meddis 1994) by hypothesizing that chopper units in the cochlear nucleus could replace the need for autocorrelation. The main attraction of this idea is the physiological implementation of a process akin to autocorrelation. The main drawback is the lack of evidence that the necessary range of units exists. The monaural  $t$ - $f$  (periodicity versus best frequency) plane hypothesized to be in the ventral cochlear nucleus (Wiegrefe and Winter 2001) is very similar to the  $t$ - $f$  plane identified by Langner and colleagues in the IC. At the level of the IC it will be important to test the hypothesis of Langner and colleagues that pitch is extracted by a series of modulation/periodicity tuned cells that lie orthogonal to the isofrequency contours. This will involve controlling for the effects of distortion and also using stimuli that are less deterministic, for example, IRN. Of course, if it isn't modulation filter banks then what is it? Alternative physiological representations of F0 at the level of the IC are conspicuous by their absence.

At the level of the auditory cortex new imaging studies are providing converging evidence that an areas beyond A1 may be involved in the coding of pitch and it will be important to test this area with relevant stimuli using animal models. Of particular concern is the failure of neurophysiologists to find cells in the auditory cortex that are representing F0. It seems equally likely that the brainstem or thalamus may contain a reasonably complete representation of many psychophysical attributes (see Nelken et al. [2003] for a more complete discussion of these issues) and that the cortex is able to modify or transform this representation by means of the numerous descending pathways that are now known to exist between the cortex and other structures. Indeed, it is known that the cortex projects as far back as the cochlear nucleus. Thus, at present, it seems more reasonable to suggest that there is a continuous interplay between ascending and descending systems.

Several topics have not been dealt with in this chapter as neurophysiologists



have very little to contribute at this point in time. It is often argued that there are two pitch mechanisms: a rate-place mechanism for resolved harmonics and a temporal mechanism for unresolved harmonics (see Plack and Oxenham, Chapter 2). As de Cheveigné points out (Chapter 6), this simply leaves us with two problems—how do we analyze the temporal information for unresolved harmonics and how do we account for the need for templates when using resolved harmonics? To date no biological implementation of templates has been found. Of course, despite several models, a neural mechanism for the extraction of F0 from predominant interspike intervals remains unproven. Frequency and F0 discrimination improve with duration but this effect is greater for unresolved harmonics. White and Plack (1998) found little improvement in discrimination performance beyond 40 ms for a resolved complex but performance improved up to 80 ms for unresolved complexes. Such time constants argue for a central, that is, supra-brainstem, role in the perception of pitch but the problem remains—how is the F0 represented in the discharges of neurons in the auditory cortex? The lack of agreement between single unit studies and new brain imaging techniques suggests that neurophysiologists have either been asking the wrong questions or looking in the wrong place. We must rely on the production of new models and/or the advent of new techniques to help in our quest for the representation of pitch in the mammalian auditory system.

*Acknowledgments.* I have had the privilege of working on the neural mechanisms of pitch perception with Roy Patterson, Lutz Wiegbebe, Daniel Pressnitzer, Jesko Verhey, and Ray Meddis at the Centre for the Neural Basis of Hearing. They, together with other members of the CNBH, have provided me with hours of thoughtful discussion and many helpful insights into the representation of pitch in the auditory system. I am grateful to the editors, Daniel Pressnitzer, Alain de Cheveigné, Lutz Wiegbebe, Veronika Neuert, and Alan Palmer for helpful comments on earlier versions of the manuscript.

## *References*

- Adams JC (1979) Ascending projections to the inferior colliculus. *J Comp Neurol* 183: 519–538.
- Adams JC (1997) Projections from octopus cells of the posteroventral cochlear nucleus to the ventral nucleus of the lateral lemniscus in cat and human. *Aud Neurosci* 3: 335–350.
- Arnott R, Wallace M, Palmer AR (2004) Onset neurons in the anteroventral cochlear nucleus project to the dorsal cochlear nucleus. *J Assoc Res. Otolaryngol* 5:153–170.
- Assmann P, Summerfield AQ (1990) Modelling the perception of concurrent vowels: vowels with different fundamental frequencies. *J Acoust Soc Am* 88:680–697.
- Biebel UW, Langner G (2002) Evidence for interactions across frequency channels in the inferior colliculus of awake chinchilla. *Hear Res* 169:151–168.

- Bilsen FA, Ritsma RJ (1969/70) Repetition pitch and its implication for hearing theory. *Acustica* 22:63–73.
- Blackburn CC, Sachs MB (1989) Classification of unit types in the anteroventral cochlear nucleus: PST histograms and regularity analysis. *J Neurophysiol* 62:1303–1329.
- Blackburn CC, Sachs MB (1990) The representation of the steady-state vowel sound /e/ in the discharge patterns of cat anteroventral cochlear nucleus neurons. *J Neurophysiol* 63:1191–1211.
- Brugge JF, Merzenich MM (1973) Patterns of activity of single neurons of the auditory cortex of monkey. In: Møller AR (ed), *Basic Mechanisms in Hearing*. New York: Academic Press, pp. 745–772.
- Cariani PA, Delgutte B (1996a) Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *J Neurophysiol* 76:1698–1716.
- Cariani PA, Delgutte B (1996b) Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase-invariance, pitch circularity, rate pitch, and the dominance region of pitch. *J Neurophysiol* 76:1717–1734.
- Carlyon RP, Shackleton TM (1994) Comparing the fundamental frequencies of resolved and unresolved harmonics: evidence for two pitch mechanisms? *J Acoust Soc Am* 95:3541–3554.
- Carlyon RP, van Wieringen A, Long CJ, Deeks JM, Wouters J (2002) Temporal pitch mechanisms in acoustic and electric hearing. *J Acoust Soc Am* 112:621–633.
- Carney L (1994) Spatiotemporal encoding of sound level: models for normal encoding and recruitment of loudness. *Hear Res* 76:31–44.
- Carney LH, Heinz MG, Evilsizer ME, Gilkey RH, Colburn HS (2002) Auditory phase opponency: a temporal model for masked detection at low frequencies. *Acta Acustica* 88:334–346.
- Casparly DM, Backoff PM, Finlayson PG, Palombi PS (1994) Inhibitory inputs modulate discharge rate within frequency receptive fields of anteroventral cochlear nucleus neurons. *J Neurophysiol* 72:2124–2133.
- Cedolin L, Delgutte B (2005) Representations of the pitch of complex tones in the auditory nerve In: Pressnitzer D, de Cheveigné, A, McAdams S, Collet L. (eds), *Auditory Signal Processing: Physiology, Psychoacoustics and Models* (in press).
- Cohen MA, Grossberg S, Wise LL (1995) A spectral network model of pitch perception. *J Acoust Soc Am* 98:862–879.
- de Cheveigné A (1993) Separation of concurrent harmonic sounds: fundamental frequency estimation and a time-domain cancellation model of auditory processing. *J Acoust Soc Am* 93:3271–3290.
- Delgutte B (1982) Some correlates of phonetic distinctions at the level of the auditory nerve. In: Granström R (ed), *The Representation of Speech in the Peripheral Auditory System*. Amsterdam: Elsevier, pp. 131–150.
- Delgutte B (1987) Peripheral auditory processing of speech information: implications from a physiological study of intensity discrimination. In: Schouten MEH (ed), *The Psychophysics of Speech Perception*. Dordrecht: Nijhoff, pp. 333–353.
- Delgutte B (1996) Physiological models for basic auditory percepts. In: Hawkins H, McMullin T, Popper AN, Fay RR (eds), *Auditory Computation*, New York: Springer-Verlag, pp. 157–220.
- Delgutte B, Kiang NYS (1984) Speech coding in the auditory nerve. I. Vowel-like sounds. *J Acoust Soc Am* 75:879–886.
- Doucet JR, Ryugo DK (1997) Projections from the ventral cochlear nucleus to the dorsal cochlear nucleus in rats. *J Comp Neurol* 385:245–264.

- Doucet, JR, Ross AT, Gillespie MB, Ryugo DK (1999) Glycine immunoreactivity of multipolar neurons in the ventral cochlear nucleus which project to the dorsal cochlear nucleus. *J Comp Neurol* 408:515–531.
- Edeline J-M (1998) Learning-induced physiological plasticity in the thalamo-cortical sensory systems: a critical evaluation of receptive field plasticity, map changes and their potential mechanisms. *Prog Neurobiol* 57:165–224.
- Eggermont JJ (2001) Between sound and perception: reviewing the search for a neural code. *Hear Res* 157:1–42.
- Ehret G, Merzenich MM (1988) Neuronal discharge rate is unsuitable for encoding sound intensity at the inferior colliculus level. *Hear Res* 35:1–18.
- Erisir A, Van Horn SC, Sherman SM (1997) Relative numbers of cortical and brainstem inputs to the lateral geniculate nucleus. *Proc Natl Acad Sci USA* 94:1517–1520.
- Evans EF (1978) Place and time coding of frequency in the peripheral auditory system: some physiological pros and cons. *Audiology* 17:369–420.
- Evans EF (1981) The dynamic range problem: Place and time coding at the level of the cochlear nerve and cochlear nucleus. In: Syka J (ed), *Neuronal Mechanisms of Hearing*. New York: Plenum Press, pp. 69–85.
- Evans EF (2001) Latest comparisons between physiological and behavioral frequency selectivity. In: Breebaart, D, Houtsma A, Kohlrausch A, Prijs V, Schoonhoven R (eds), *Proceedings of the 12th International Symposium on Hearing, Physiological and Psychophysical Bases of Auditory Function*. Maastricht: Shaker BV, pp. 382–387.
- Evans EF, Palmer AR (1980) Relationship between the dynamic ranges of cochlear nerve fibers and their spontaneous activity *Exp Brain Res* 40:115–118.
- Evans EF, Zhao W (1998) Periodicity coding of the fundamental frequency of harmonic complexes: physiological and pharmacological study of onset units in the ventral cochlear nucleus. In: Palmer AR, Rees A, Summerfield AQ, Meddis R (eds), *Psychophysical and Physiological Advances in Hearing*. London: Whurr, pp. 186–194.
- Fishman YI, Reser DH, Arezzo JC, Steinschneider M (1998) Pitch vs. spectral encoding of harmonic complex tones in primary auditory cortex of the awake monkey. *Brain Res* 786:18–30.
- Frisina RD, Smith RL, Chamberlain SC (1990). Encoding of amplitude modulation in the gerbil cochlear nucleus: I. A hierarchy of enhancement. *Hear Res* 44:99–122.
- Frisina RD, Walton JP, Karcich KJ (1994) Dorsal cochlear nucleus single neurons can enhance temporal processing capabilities in background noise. *Exp Brain Res* 102:160–164.
- Frisina RD, Karich KJ, Tracy TC, Sullivan DM, Walton JP, Colombo J (1996) Preservation of amplitude modulation coding in the presence of background noise by chinchilla auditory-nerve fibers. *J Acoust Soc Am* 99:475–490.
- Fritz J, Shamma S, Elhilali M, Klein D (2003) Rapi-task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat Neurosci* 6:1216–1223.
- Geisler CD, Silkes SM (1991) Responses of “lower-spontaneous rate” auditory nerve fibers to speech syllables presented in noise. II. Glottal-pulse periodicities. *J Acoust Soc Am* 90:3140–3148.
- Godfrey, DA Kiang NYS, Norris BE (1975) Single unit activity in the posteroventral cochlear nucleus of the cat *J Comp Neurol* 162:247–268.
- Goldberg J, Brown PB (1969) Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localisation. *J Neurophysiol* 32:613–636.

- Griffiths TD, Buchel C, Frackowiak RSJ, Patterson RD (1998) Analysis of temporal structure in sound by the human brain. *Nat Neurosci* 1:422–427.
- Heil P, Rajan R, Irvine DRF (1994) Topographic representation of tone intensity along the iso-frequency axis of cat primary auditory cortex. *Hear Res* 76:188–202.
- Heinz MG, Colburn HS, Carney LH (2001) Evaluating auditory performance limits: I. One parameter discrimination using a computational model for the auditory nerve. *Neural Comput* 13:2273–2316.
- Hewitt MJ, Meddis R (1994) A computer model of amplitude modulation sensitivity of single units in the inferior colliculus. *J Acoust Soc Am* 95:2145–2159.
- Horst JW, Javel E, Farley GR (1985) Extraction and enhancement of spectral structure by the cochlea. *J Acoust Soc Am* 78:1898–1901.
- Horst JW, Javel E, Farley GR (1986) Coding of spectral fine structure in the auditory nerve. I. Fourier analysis of period and interspike interval histograms. *J Acoust Soc Am* 79:398–416.
- Horst JW, Javel E, Farley GR (1990) Coding of spectral fine structure in the auditory nerve. II. Level dependent nonlinear responses. *J Acoust Soc Am* 88:2656–2681.
- Hose B, Langner G, Scheich H (1987) Topographic representation of periodicities in the forebrain of the Mynah bird: one map for pitch and rhythm? *Brain Res* 422:367–373.
- Irvine DRF, Gago (1990) Binaural interaction in high-frequency neurons in inferior colliculus of the cat: effects of variations in sound pressure level on sensitivity to interaural intensity differences. *J Neurophysiol* 63:570–591.
- Javel E (1980) Coding of AM tones in the Chinchilla auditory nerve: implications for the pitch of complex tones. *J Acoust Soc Am* 68:133–146.
- Johnson D (1980) The relationship between spike rate and synchrony in responses of auditory nerve fibers to single tones. *J Acoust Soc Am* 68:1115–1122.
- Joris P, Smith PH (1998) Temporal and binaural properties in dorsal cochlear nucleus and its output tract. *J Neurosci* 18:10157–10170.
- Joris PX, Carney LH, Smith PH, Yin TC (1994) Enhancement of neural synchronization in the anteroventral cochlear nucleus. I. Responses to tones at the characteristic frequency. *J Neurophysiol* 71:1022–1036.
- Joris P, Schreiner CE, Rees A (2004) Neural processing of amplitude-modulated sounds. *Physiol Rev* 84:541–577.
- Kadia SC, Wang X (2003) Spectral integration in A1 of awake primates: neurons with single- and multi-peaked tuning characteristics. *J Neurophysiol* 89:1603–1622.
- Kaernbach C, Demany L (1998) Psychophysical evidence against the autocorrelation theory of pitch perception. *J Acoust Soc Am* 104:2298–2306.
- Keilson SE, Richards VM, Wyman BT Young ED (1997) The representation of concurrent vowels in the cat anaesthetized ventral cochlear nucleus: evidence for a periodicity-tagged spectral representation. *J Acoust Soc Am* 102:1056–1071.
- Kilgard MO, Merzenich MM (1998) Cortical map reorganization enabled by nucleus basalis activity. *Science* 279:1714–1718.
- Kim DO, Leonard G (1988) Pitch-period following response of cat cochlear nucleus neurons to speech sounds. In: Duifhuis H, Horst JW, Wit HP (eds), *Basic Issues in Hearing*, London: Academic Press, pp. 252–260.
- Kim DO, Molnar CE (1979) A population study of cochlear nerve fibers: comparison of spatial distributions of average-rate and phase locking measures of responses to single tones. *J Neurophysiol* 42:16–30.
- Kim DO, Parham K (1990) Auditory nerve spatial encoding of high frequency pure tones:

- population response profiles derived from  $d'$  measure associated with nearby places along the cochlea. *Hear Res* 52:167–180.
- Kim DO, Rhode WS, Greenberg, SR (1986) Responses of cochlear nucleus neurons to speech signals: neural encoding of pitch, intensity and other parameters. In: Moore BCJ, Patterson RD (eds), *Auditory Frequency Selectivity: A NATO Advanced Research Workshop*. New York: Plenum Press, pp. 281–288.
- Kim DO, Chang SO, Sirianni JG (1990a) A population study of auditory nerve fibers in unanaesthetized decerebrate cats: responses to pure tones. *J Acoust Soc Am* 87:1648–1655.
- Kim DO, Sirianni, JG, Chang SO (1990b) Responses of DCN-PVCN neurons and auditory-nerve fibers in unanaesthetized decerebrate cats to AM and pure tones: analysis with autocorrelation/power spectrum. *Hear Res* 45:95–113.
- Kim DO, Parham K, Sirianni JG, Chang, SO (1991) Spatial response profiles of posteroventral cochlear nucleus neurons and auditory nerve fibers in unanaesthetized decerebrate cats: responses to pure tones. *J Acoust Soc Am* 89:2804–2817.
- Kopp-Scheinpflug C, Dehmel S, Dorrscheidt GJ, Rubsamen R (2002) Interaction of excitation and inhibition in anteroventral cochlear nucleus neurons that receive large endbulb synaptic endings. *J Neurosci* 22:11004–11018.
- Koppl C (1997) Phase locking to high frequencies in the auditory nerve and cochlear nucleus magnocellularis of the Barn Owl, Tyto. *Alba J Neurosci* 17:3312–3321.
- Krishna BS, Semple MN (2000) Auditory temporal processing: responses to sinusoidally amplitude modulated tones in the inferior colliculus. *J Neurophysiol* 84:255–273.
- Krumbholz K, Patterson RD, Pressnitzer D (2000) The lower limit of pitch as determined by rate discrimination. *J Acoust Soc Am* 108:1170–1180.
- Lai Y-C, Winslow RL, Sachs MB (1994) The functional role of excitatory and inhibitory interactions in chopper cells of the anteroventral cochlear nucleus. *Neural Comput* 6: 1127–1140.
- Langner G (1981) Neuronal mechanisms for pitch analysis in the time domain. *Exp Brain Res* 44:450–454.
- Langner G (1988) Physiological properties of units in the cochlear nucleus are adequate for a model of periodicity analysis in the auditory midbrain. In: Syka J, Masterton RB (eds), *Auditory Pathway*. New York: Plenum Press, pp. 207–212.
- Langner G, Schreiner CE (1988) Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. *J Neurophysiol* 60:1799–1822.
- Langner G, Albert M, Briede T (2002) Temporal and spatial coding of periodicity information in the inferior colliculus of the awake chinchilla (*Chinchilla laniger*). *Hear Res* 168:110–130.
- Liberman MC (1978) Auditory-nerve response from cats raised in a low noise chamber. *J Acoust Soc Am* 63:442–455.
- Liberman MC (1982) Single-neuron labeling in the cat auditory nerve. *Science* 216: 1239–1241.
- Liberman MC (1988) Physiology of cochlear efferent and afferent neurons: direct comparisons in the same animal. *Hear Res* 34:179–192.
- Liberman MC (1991) Central projections of auditory nerve fibers of differing spontaneous rate. I. Anteroventral cochlear nucleus. *J Comp Neurol* 313:240–258.
- Liberman MC (1993) Central projections of auditory nerve fibers of differing spontaneous rate, II: posteroventral and dorsal cochlear nuclei. *J Comp Neurol* 327:17–36.

- Lieberman MC, Oliver ME (1984) Morphometry of intracellularly labeled neurons of the auditory nerve: correlations with functional properties. *J Comp Neurol* 223:163–176.
- Lu T, Liang L, Wang X (2001) Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. *Nat Neurosci* 4:1131–1138.
- May BJ, Sachs MB (1992) Dynamic range of neural rate responses in the ventral cochlear nucleus of awake cats. *J Neurophysiol* 68:1589–1602.
- May BJ, Huang A, Le Prell G, Heinz RD (1996) Vowel formant frequency discrimination in cats: comparison of auditory nerve representations and psychophysical thresholds. *Audit Neurosci* 3:135–162.
- May BJ, Le Prell GS, Sachs MB (1998) Vowel representations in the ventral cochlear nucleus of the cat: effects of level, background noise and behavioral state. *J Neurophysiol* 79:1755–1767.
- McAlpine DM (2004) Neural sensitivity to periodicity in the inferior colliculus: Evidence for the role of cochlear distortions. *J Neurophysiol* 92:1295–1311.
- McAlpine DM, Grothe B (2003) Sound localisation and delay lines—do mammals fit the model? *Trends Neurosci* 26:347–350.
- Meddis R, Hewitt MJ (1991a) Virtual pitch and phase sensitivity studied using a computer model of the auditory periphery. I: Pitch identification. *J Acoust Soc Am* 89:2866–2882.
- Meddis R, Hewitt MJ (1991b) Virtual pitch and phase sensitivity studied using a computer model of the auditory periphery. II: Phase sensitivity *J Acoust Soc Am* 89:2883–2894.
- Merzenich, MM, Knight PL, Roth, GL (1975) Representation of the cochlea within primary auditory cortex in the cat. *J Neurophysiol* 38:231–249.
- Miller MI, Sachs MB (1984) Representation of voice pitch in discharge patterns of auditory-nerve fibers. *Hear Res* 14:257–279.
- Nelken I, Young ED (1994) Two separate inhibitory mechanisms shape the responses of dorsal cochlear nucleus type IV units to narrowband and wideband stimuli. *J Neurophysiol* 71:2446–2462.
- Nelken I, Fishbach A, Las L, Ulanovsky N, Farkas D (2003) Primary auditory cortex of cats: feature detection or something else? *Biol Cybernetics* 89:397–406.
- Oertel D, Wu SH, Garb MW, Dizack C (1990) Morphology and physiology of cells in slice preparations of the posteroventral cochlear nucleus of mice. *J Comp Neurol* 295:136–154.
- Oertel D, Bal R, Gardner SM, Smith PH, Joris PX (2000) Detection of synchrony in the activity of auditory nerve fibers by octopus cells of the mammalian cochlear nucleus. *Proc Natl Acad Sci USA* 97:11773–11779.
- Oliver DL, Morest DK (1984) The central nucleus of the inferior colliculus in the cat. *J Comp Neurol* 222:237–264.
- Oxenham AJ, Shera CA (2003) Estimates of human cochlear tuning at low levels using forward and simultaneous masking. *J Assoc Res Otolaryngol* 4:541–554.
- Palmer AR (1990) The representation of the spectra and fundamental frequencies of steady-state single- and double-vowel sounds in the temporal discharge patterns of guinea pig cochlear nerve fibers. *J Acoust Soc Am* 88:1412–1426.
- Palmer AR, Russell IJ (1986) Phase locking in the cochlear nerve of the guinea pig and its relation to the receptor potential of inner hair cells. *Hear Res* 24:1–15.
- Palmer AR, Winter IM (1992) Cochlear nerve and cochlear nucleus response to the fundamental frequency of voiced speech sounds and harmonic complex tones. In: Cazals Y, Demany L, Horner K (eds), *Auditory Physiology and Perception*. Oxford: Pergamon, pp. 231–239.

- Palmer AR, Winter IM (1993) Coding of the fundamental frequency of voiced speech sounds and harmonic complex tones in the ventral cochlear nucleus. In: Merchan MA, Juiz J, Godfrey DA, Mugnaini E (eds), *Mammalian Cochlear Nuclei: Organization and Function*. New York: Plenum Press, pp. 373–384.
- Palmer AR, Winter IM (1996) The temporal window of two-tone facilitation in onset units of the ventral cochlear nucleus. *Audiol Neurotol* 1:12–30.
- Palmer AR, Winter IM, Darwin CJ (1986) The representation of steady-state vowels in the temporal discharge patterns of the guinea pig cochlear nerve and primarylike cochlear nucleus neurons. *J Acoust Soc Am* 79:100–113.
- Palmer AR, Jiang D, Marshall D (1996) Responses of ventral cochlear nucleus onset and chopper units as a function of signal bandwidth. *J Neurophysiol* 75:780–794.
- Pantev C, Hoke M, Lutkenhoner B, Lehnertz K (1989) Tonotopic organization of the auditory cortex: pitch versus frequency representation *Science* 246:486–488.
- Patterson RD (1994) The sound of a sinusoid: spectral models. *J Acoust Soc Am* 96:1409–1418.
- Patterson RD, Allerhand MH, Giguere C. (1995) Time-domain modeling of peripheral auditory processing: a modular architecture and a software platform. *J Acoust Soc Am* 98:1890–1894.
- Patterson RD, Uppenkamp S, Johnsrude I, Griffiths TD (2002) The processing of temporal pitch and melody information in auditory cortex. *Neuron* 36:767–776.
- Pfeiffer RR, Kim DO (1975) Cochlear nerve fiber responses distribution along the cochlear partition. *J Acoust Soc Am* 58:867–869.
- Phillips DP, Orman SS (1984) responses of single neurons in posterior field of cat auditory cortex to tonal stimulation. *J Neurophysiol* 51:147–163.
- Phillips DP, Orman SS, Musicant AD, Wilson GF (1985) Neurons in the cat's primary auditory cortex distinguished by their responses to tones and wide spectrum noise. *Hear Res* 18:73–86.
- Phillips DP, Semple MN, Calford MB, Kitzes LM (1994) Level-dependent representation of stimulus frequency in cat primary auditory cortex. *Exp Brain Res* 102:210–226.
- Pressnitzer D, Patterson RD (2001) Distortion products and the pitch of harmonic complex tones. In: Breebaart D, Houtsma A, Kohlrausch A, Prijs V, Schoonhoven R (eds), *Proceedings of the 12th International Symposium on Hearing, Physiological and Psychophysical Bases of Auditory Function*. Maastricht: Shaker BV, pp. 97–104.
- Pressnitzer D, de Cheveigné A, Winter IM (2001) Perceptual pitch shift for sounds with similar waveform autocorrelation. *Acoust Res Lett Online* 3:1–6.
- Pressnitzer D, de Cheveigné A, Winter IM (2004) Physiological correlates of the perceptual pitch shift for sounds with similar waveform autocorrelation. *Acoust Res Lett Online* 5:1–6.
- Rees A, Palmer AR (1988) Rate-intensity functions and their modification by broadband noise. *J Acoust Soc Am* 83:1488–1498.
- Rhode WS (1994) Temporal encoding of 200% amplitude modulated signals in the ventral cochlear nucleus of the cat. *Hear Res* 77:43–68.
- Rhode WS (1995) Interspike intervals as a correlate of periodicity in cat cochlear nucleus. *J Acoust Soc Am* 97:2414–2429.
- Rhode WS, Smith PH (1986) Encoding timing and intensity in the ventral cochlear nucleus of the cat. *J Neurophysiol* 56:261–286.
- Robertson D, Irvine DRF (1989) Plasticity of frequency organization in auditory cortex of guinea pigs with partial unilateral deafness. *J Comp Neurol* 282:456–471.
- Robles L, Ruggero MA (2001) Mechanics of the mammalian cochlea. *Physiol Rev* 81:1305–1352.

- Rockel AJ, Jones EG (1973a) The neuronal organization of the inferior colliculus of the adult cat. I. The central nucleus. *J Comp Neurol* 147:11–60.
- Rockel AJ, Jones EG (1973b) Observations on the fine structure of the central nucleus of the inferior colliculus of the cat. *J Comp Neurol* 147:61–92.
- Rose JE, Galambos R, Hughes JR (1959) Microelectrode studies of the cochlear nuclei of the cat. *Bull John Hopkins Hosp* 104:211–251.
- Sachs MB, Abbas PJ (1974) Rate-versus level functions for auditory-nerve fibers in cats: tone-burst stimuli. *J Acoust Soc Am* 56:1835–1847.
- Sachs MB, Young ED (1979) Encoding of steady-state vowels in the auditory nerve: representation in terms of discharge rate. *J Acoust Soc Am* 66:470–479.
- Schofield BR (1995) Projections from the cochlear nucleus to the superior para-olivary nucleus in guinea pigs. *J Comp Neurol* 360:135–149.
- Schofield BR, Cant NB (1997) Ventral nucleus of the lateral lemniscus in guinea pigs: cytoarchitecture and inputs from the cochlear nucleus. *J Comp Neurol* 379:363–385.
- Schouten J (1940) The residue and the mechanism of hearing. *Proc K Ned Akad Wet* 43:991–999.
- Schreiner CE, Langner G (1988) Periodicity coding in the inferior colliculus of the cat II. Topographical organisation. *J Neurophysiol* 60:1823–1840.
- Schrottge I, Scheich H, Schuze H (2004) Neuronal responses to amplitude modulated sounds in the Mongolian gerbil auditory midbrain and cortex: periodicity coding or responses to distortion products? *Assoc Res Otolaryngol Abstr* 27:289.
- Schulze H, Langner G (1999) Auditory cortical responses to amplitude modulations with spectra above frequency receptive fields: evidence for wide spectral integration. *J Comp Physiol* 185:493–508.
- Schulze H, Hess A, Ohl FW, Scheich H (2002) Superposition of horseshoe-like periodicity and linear tonotopic maps in auditory cortex of the Mongolian gerbil. *Eur J Neurosci* 15:1077–1084.
- Schwartz DWF, Tomlinson RWW (1990) Spectral response properties of auditory cortex neurons to harmonic complex tones in alert monkey (*Macaca mulatta*). *J Neurophysiol* 64:282–299.
- Semal C, Demany L (1990) The upper limit of musical pitch. *Music Percept* 8:165–176.
- Semple MN, Kitzes LM (1987) Binaural processing of sound pressure level in the inferior colliculus. *J Neurophysiol* 57:1130–1147.
- Semple MN, Scott BH (2003) Cortical mechanisms in hearing. *Curr Opin Neurobiol* 13:167–173.
- Shamma SA (1985a) Speech processing in the auditory system. I. The representation of speech sound in the responses of the auditory nerve. *J Acoust Soc Am* 78:1612–1621.
- Shamma SA (1985b) Speech processing in the auditory system. II. Lateral inhibition and the central processing of speech evoked activity in the auditory nerve. *J Acoust Soc Am* 78:1622–1632.
- Shamma SA, Klein D (2000) The case of the missing pitch templates: how harmonic templates emerge in the early auditory system. *J. Acoust Soc Am* 107:2631–2644.
- Shera CA, Guinan JJ, Oxenham AJ (2002) Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. *Proc Natl Acad Sci USA* 99:3318–3323.
- Shofner WP (1991) Temporal representation of rippled noise in the anteroventral cochlear nucleus of the chinchilla. *J Acoust Soc Am* 90:2450–2466.



- Shofner WP (1999) Responses of cochlear nucleus units in the chinchilla to iterated rippled noises: quantitative analysis of neural autocorrelograms of primarylike and chopper units. *J Neurophysiol* 81:2662–2674.
- Shofner WP, Dye R (1989) Statistical and receiver operating characteristic analysis of empirical spike count distributions: quantifying the ability of cochlear nucleus units to signal intensity changes. *J Acoust Soc Am* 86:2171–2184.
- Shofner WP, Sachs, MB (1986) Representation of a low-frequency tone in the discharge rate of populations of auditory nerve fibers. *Hear Res* 21:91–95.
- Siddhartha KC, Wang X (2003) Spectral integration in A1 of awake primates: neurons with single and multi-peaked tuning characteristics. *J Neurophysiol* 89:1603–1622.
- Slaney M, Lyon R (1990) A perceptual pitch detector. *Proc ICASSP 90*, Albuquerque, New Mexico.
- Smith PH, Rhode WS (1989) Structural and functional properties distinguish two types of multipolar cells in the ventral cochlear nucleus. *J Comp Neurol* 282:595–616.
- Smith PH, Joris PX, Banks MI, Yin TCT (1993) Responses of cochlear nucleus cells and projections of their axons. In: Merchan MA, Juiz JM, Godfrey DA, Mugnaini E (ed), *The Mammalian Cochlear Nuclei: Organisation and Function*. New York: Plenum Press, pp. 349–360.
- Srulovicz P, Goldstein JL (1983) A central spectrum model: synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum. *J Acoust Soc Am* 73:1266–1276.
- Steinschneider M, Reser DH, Fishnan YI, Schroeder CE, Arezzo JC (1998) Click train encoding in primary auditory cortex of the awake monkey: evidence for two mechanisms subserving pitch perception *J Acoust Soc Am* 104:2935–2955.
- Suga N, Gao E, Zhang Y, Ma X, Olsen JF (2000) The corticofugal system for hearing: recent progress. *Proc Natl Acad Sci USA* 97:11807–11814.
- Sutter ML, Schreiner CE (1991) Physiology and topography of neurons with multi-peaked tuning curves in cat primary auditory cortex. *J Neurophysiol* 65:1207–1226.
- Terhardt E (1975) Influence of intensity on the pitch perception of complex tones. *Acustica* 33:344–348.
- Tomlinson RWW, Schwartz DWF (1988) Perception of the missing fundamental in non-human primates. *J Acoust Soc Am* 84:560–565.
- Ulanovsky N, Las L, Nelken I (2003) Processing of low-probability sounds by cortical neurons. *Nat Neurosci* 6:391–398.
- Vater M, Covey E, Casseday, JH (1997) The columnar region of the ventral nucleus of the lateral lemniscus in the big brown bat (*Eptesicus fuscus*): synaptic arrangements and structural correlates of feedforward inhibitory function. *Cell Tissue Res* 289:223–233.
- Verhey JL, Neuert V, Winter IM (2004) Responses of single units in the mammalian cochlear nucleus to iterated rippled noise with negative gain. *Assoc Res Otolaryngol Abstr* 27:309.
- Wallace MN, Shackleton TM, Palmer AR (2002) Phase-locked responses to pure tones in the primary auditory cortex. *Hear Res* 163:1–12.
- Weiss TF, Rose C (1988) A comparison of synchronization filters in different auditory receptor organs. *Hear Res* 33:175–180.
- Wessinger CM, VanMeter J, Tian B, Van Lare J, Pekar J, Rauschecker JP (2001) Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *J Cogn Neurosci* 13:1–7.

- White LJ Plack CJ (1988) Temporal processing of the pitch of complex tones. *J Acoust Soc Am* 103:2051–2063.
- Whitfield IC (1980) Auditory cortex and the pitch of complex tones. *J Acoust Soc Am* 67:644–647.
- Wiegerebe L, Meddis R (2004) The representation of periodic sounds in simulated sustained chopper units of the ventral cochlear nucleus. *J Acoust Soc Am* 115:1207–1218.
- Wiegerebe L, Patterson RD (1999) The role of modulation in the pitch of high-pass filtered iterated rippled noise. *Hear Res* 132:94–108.
- Wiegerebe L, Winter IM (2001) Temporal representation of iterated rippled noise as a function of delay and sound level in the ventral cochlear nucleus. *J Neurophysiol* 85:1206–1219.
- Willard FH, Ryugo DK (1983) Anatomy of the central auditory system. In: Willot JF (ed), *The Auditory Psychobiology of the Mouse*. Springfield, IL: Charles C. Thomas, pp. 201–304.
- Winslow R, Sachs MB (1988) Single tone intensity discrimination based on auditory nerve fiber responses in backgrounds of quiet, noise and with stimulation of the crossed olivocochlear bundle. *Hear Res* 35:165–190.
- Winter IM, Palmer AR (1990a) Responses of single units in the anteroventral cochlear nucleus of the guinea pig. *Hear Res* 44:161–178.
- Winter IM, Palmer AR (1990b) Temporal responses of primarylike anteroventral cochlear nucleus units to the steady-state vowel /i/. *J Acoust Soc Am* 88:1437–1441.
- Winter IM, Palmer AR (1995) Level dependence of cochlear nucleus onset unit responses and facilitation by second tones or broadband noise. *J Neurophysiol* 73:141–159.
- Winter IM, Robertson D, Yates GK (1990) Diversity of characteristic frequency rate-level functions in guinea pig auditory nerve fibers. *Hear Res* 45:191–202.
- Winter IM, Wiegerebe L, Patterson RD (2001) The temporal representation of the delay of iterated rippled noise in the ventral cochlear nucleus of the guinea pig. *J Physiol* 537:553–566.
- Yan J, Ehret G (2002) Corticofugal modulation of midbrain sound processing in the house mouse. *Eur J Neurosci* 16:119–128.
- Yates GK, Robertson D, Winter IM (1990) Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range. *Hear Res* 45:203–220.
- Yost WA, Patterson RD, Sheft S (1996) A time domain description for the pitch strength of iterated rippled noise. *J Acoust Soc Am* 99:1066–1078.
- Young ED, Barta P (1986) Rate responses of auditory nerve fibres to tones in noise near masked threshold. *J Acoust Soc Am* 79:426–442.
- Young ED, Sachs MB (1979) Representation of steady-state vowels in the temporal aspects of discharge patterns of populations of auditory nerve fibers. *J Acoust Soc Am* 66:1381–1403.
- Young ED, Sachs MB (1980) Effects of nonlinearities on speech coding in the auditory nerve. *J Acoust Soc Am* 68:858–875.

# Functional Imaging of Pitch Processing

TIMOTHY D. GRIFFITHS

## 1. Introduction

This chapter considers the application of brain imaging techniques to address two questions related to pitch perception. The first question is: How does the brain process stimulus properties that are relevant to the perception of pitch? This book is primarily about the percept called pitch rather than the representation of auditory stimuli, and the second, more difficult, question relates to whether the imaging techniques allow any comment on the neural correlates of this percept. Functional imaging is used here to refer to both the hemodynamic techniques—positron emission tomography (PET) and functional magnetic resonance imaging (fMRI)—and the electromagnetic techniques—electroencephalography (EEG) and magnetoencephalography (MEG). The hemodynamic and electromagnetic techniques will be considered separately, but should be regarded as complementary methods with different strengths and weaknesses. The hemodynamic techniques are based on the imaging of signals related to regional blood flow; they allow a measurement of activity in the whole brain with a spatial precision that can be less than 1 cm, but cannot be used to follow brain activity in the form of rapid temporal patterns where brain activity changes occur over a time scale of less than a second. Electromagnetic techniques allow the measurement of electrical changes in the brain with millisecond accuracy, but require a number of assumptions to map the origin of such activity.

## 2. Hemodynamic Techniques: PET and fMRI

### *2.1 Background: The Basis for (and Limitations of) Hemodynamic Techniques*

This section considers how hemodynamic imaging can contribute to our understanding of pitch processing in humans. Functional imaging is used here to refer to PET and fMRI, which are both techniques that depend on the blood

flow response to brain activity. In PET, regional cerebral blood flow is measured directly using a radioactive tracer, while in fMRI the blood oxygenation level dependent (BOLD) response is measured. It is only recently that direct evidence has demonstrated a direct link between the hemodynamic response and the local brain activity. Logothetis et al. (2001) measured both the hemodynamic response using BOLD and the local neural brain activity in the macaque in response to a visual stimulus. The local brain activity was assessed both by the local field activity, a measure of dendritic activity, and the multi-unit activity, a measure of axonal activity. This important work represents the first direct demonstration of the link between the BOLD response and neuronal activity. The best correlation was found between BOLD and the local field potential, suggesting that BOLD reflects the dendritic input to neurons rather than their axonal output. Recent work suggests a particular importance of glial cells in this coupling (Parri and Crunelli 2003). The correlation of BOLD and dendritic input is worth bearing in mind when considering the interpretation of functional imaging experiments. Activation in a given area during a particular aspect of pitch processing may reflect dendritic activity in response to inputs from local neurons resulting from the extensive vertical connections in cortical areas. But it could also occur, in principle, due to dendritic activity in response to input from other subcortical or cortical areas. In other words, the location of the neuronal cell type that primarily responds to a given type of stimulus and the location of the resultant hemodynamic response could be different.

Much debate in pitch processing centers on the relative importance of different types of neural codes, and it is important to realize what type of coding can be demonstrated using functional imaging techniques. The hemodynamic response is slow, typically of the order of 10 s in cortex (Hall et al. 1999). Recent work has sought to identify transient and sustained components of the hemodynamic response in auditory cortex in response to sound (Seifritz et al. 2002) but even for the transient response the onset time (time to 10% peak) is approximately 3 s. These responses can therefore only reflect the integrated activity over what is (in neurophysiological terms) a very long time window. The Logothetis et al. (2001) work confirms that the hemodynamic response is related to the mean local firing rate, a population rate code. Temporal encoding corresponding to pitch processing cannot be directly assessed using such measures. In a number of studies carried out by our group (Griffiths et al. 1998, 2001; Patterson et al. 2002) the temporal regularity in sounds has been manipulated and the resulting change in the hemodynamic response assessed. Based on the preceding arguments, these studies do not demonstrate temporal codes in the brain; rather, they show changes in the local mean firing rate in response to the changes in temporal regularity. These studies therefore represent a test of models of temporal encoding where the regularity of temporal firing patterns is converted to a more stable population rate code.

Another critical question when considering pitch processing and functional imaging is whether the hemodynamic responses that we observe correspond to the encoding of stimulus properties, or whether they correspond to neural cor-

relates of the conscious perception (Frith et al. 1999) that is called pitch. In many experiments it is very difficult to tell. For example, in the experiments where temporal regularity is varied, there is no absolute way of interpreting the neural activity that is measured as a correlate of the stimulus properties, or as a correlate of the percept that is generated. All that can be said is that the mean local firing rate increases in certain areas in response to the stimulus manipulation. Auditory neuroscience is a little behind visual neuroscience in this respect. In visual neuroscience, for example, a number of functional imaging experiments have looked at the brain response to bistable percepts such as binocular rivalry (e.g., Lumer et al. 1998), in which fixed stimulus properties can lead to a varying percept. These influential studies allow inference about the neural correlates of perception. In the case of pitch, the development of such stimuli for imaging experiments could lead to important insights in the future. Certain experiments using complex pitch (e.g., that associated with the missing fundamental) could be interpreted as showing a mapping of the percept of pitch rather than stimulus properties. However, these experiments can also be interpreted in terms of a mapping of the stimulus property of temporal envelope.

## *2.2 The Processing of Stimulus Properties Relevant to Pitch in the Ascending Auditory Pathway*

Recent experiments have identified the response of structures in the ascending auditory pathway to systematic variation in the regularity of the stimulus (Griffiths et al. 2001). This work was made possible by two recent advances in auditory functional imaging. The first advance is the use of sparse imaging (Hall et al. 1999) to overcome the considerable noise that is produced by the MRI scanner during the measurement of the BOLD response (Ravicz et al. 2000). Sparse imaging uses the sluggishness of the BOLD response to advantage. Essentially, infrequent scans are carried out at intervals on the order of 10 s. Because of the slow buildup of the BOLD response (typically 10 s to peak in primary auditory cortex), the brain activity measured in this way will correspond to the period before scanning when the test stimuli were presented in quiet (without “contaminating” scanner noise). Apart from the improvement in measurement of the brain signal, there are good biological reasons to measure the brain responses without the additional noise produced during scanning. The presence of the additional noise will change the nature of the listening task into one for which the substrate may not be the same as the detection of the stimulus in silence. The disadvantage of sparse imaging (as opposed to “epoch mode” scanning where more frequent scans are carried out) is that it can take considerable time for sufficient scans to be carried out to allow statistical analysis of the effect of the stimulus on the brain signal.

Sparse imaging produces an improvement in the BOLD signal-to-noise ratio in both the brainstem and cortex. A second recent advance, cardiac triggering, is particularly relevant to brainstem imaging. This technique was pioneered by Melcher and colleagues, who used a limited number of slices passing through

brainstem structures of interest (Guimares et al. 1998). Cardiac triggering is a way of overcoming the considerable movement of the brainstem caused by pulsation of the basilar artery. Most stimulus analysis software, including Statistical Parametric Mapping (<http://www.fil.ion.ucl.ac.uk/spm>), incorporates algorithms that can correct for movement. However, in the brainstem there is a degradation of the BOLD response due to movement that cannot be corrected for. In cardiac triggering, the start of each scan is triggered by the R wave of the electrocardiogram, so that the scan always occurs at the same point in the cardiac cycle when the brainstem is in the same place. This technique has been used with an ascending axial acquisition in which the lower slices through the brainstem are acquired first. Figure 5.1 shows the images obtained, which are the first images to show simultaneous sound activation in the whole of the human auditory brain. The sections show areas where there is significant activation for the comparison between a sound stimulus and a silent condition. Activation can be seen in the cochlear nucleus (CN), lateral lemniscus (LL), inferior colliculus (IC), medial geniculate body (MGB), and auditory cortex. The peak activation that occurred was compared with the location of the structures of the ascending pathway

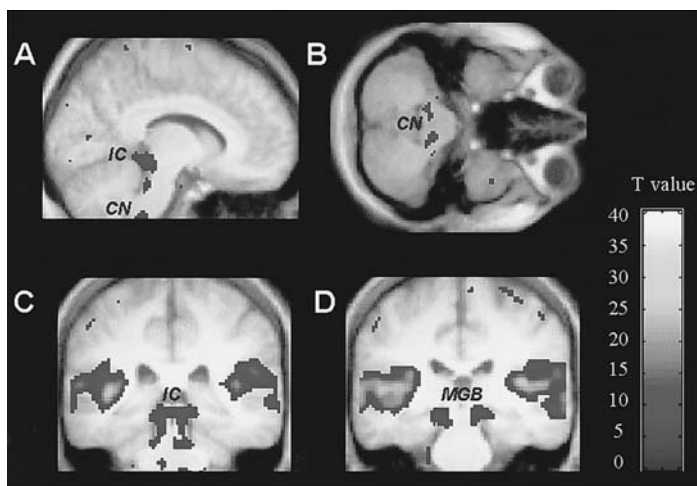


FIGURE 5.1. fMRI BOLD activation of structures in the ascending auditory pathway with sound stimuli using cardiac triggering and sparse imaging (contrast between all sound stimuli and silence shown in relation to average structural MRI). (A) Sagittal at  $x = 10$  mm showing activation in the right cochlear nucleus and inferior colliculus. (B) Axial section at  $z = -46$  mm showing bilateral activation of cochlear nuclei. (C) Coronal section at  $y = -34$  mm showing activation of inferior colliculi and superior temporal cortex. (D) Coronal section at  $y = -28$  mm showing activation of medial geniculate bodies. Threshold for contrast  $p < 0.001$  (uncorrected). Color scale gives Student's  $t$  statistic for the comparison between the BOLD values in the sound conditions and rest. (See color insert.) Reproduced from Griffiths et al. (2001), with permission, © Nature Publishing Group.

identified for each subject using the anatomical MRI scans (Table 5.1). From Table 5.1 it can be seen that there is very good correspondence between the structurally defined centers and the functional activation.

The sound stimuli used in this study were regular interval sounds in the form of iterated rippled noise (Yost et al. 1996). These noises are created by using a delay-and-add algorithm that produces regularity in the stimulus and a pitch. The strength of the pitch corresponds to the regularity of the stimulus (measured by the height of the first peak in the autocorrelation function; see also Plack and Oxenham, Chapter 2). In these experiments the pitch of the sound is kept low (50 to 100 Hz) and the sounds are high-pass filtered at 500 Hz to minimize the resolvable spectral change (the ripple in the spectrum as represented in the auditory nerve) due to the delay-and-add process. Under these conditions, changes to the stimulus and its auditory representation in the time domain are the most parsimonious explanation for the pitch that is perceived.

The temporal regularity of the stimulus was varied by changing the number of iterations in the delay-and-add process. A volume-of-interest analysis was carried out on each of the structures of the ascending auditory pathway to test the hypothesis that there is a relationship between the local brain activity, mea-

TABLE 5.1. Coordinates of structures in the ascending pathway determined from individual structural data on eight subjects and from the group sound-minus-silence comparison, with significance levels for the contrasts between sound conditions.

| Structure | Mean structural coordinates [SD] | Functional coordinates (group, sound-minus-silence contrast) |
|-----------|----------------------------------|--------------------------------------------------------------|
| CN-l      | -8 -41 -47<br>[1.3 1.1 1.8]      | -12 -40 -46                                                  |
| CN-r      | 9 -41 -48<br>[1.1 1.1 1.6]       | 8 -34 -48                                                    |
| IC-l      | -5 -35 -10<br>[0.5 0.9 1.3]      | -6 -34 -12                                                   |
| IC-r      | 5 -35 -10<br>[0.7 0.9 1.3]       | 6 -36 -10                                                    |
| MGB-l     | -16 -26 -8<br>[1.4 1.8 1.9]      | -16 -28 -10                                                  |
| MGB-r     | 16 -26 -6<br>[1.1 2.3 2.2]       | 10 -32 -8                                                    |

The coordinates are in millimeters in Talairach space (Talairach and Tournoux 1988). Mean structural coordinates refers to the mean of the coordinates that were determined for each subject using a structural algorithm. Significance levels are given after correction for multiple comparisons within the volume-of-interest defined. CN, cochlear nucleus; IC, inferior colliculus; MGB, medial geniculate body. Modified from Griffiths et al. (2001), with permission, © Nature Publishing Group.

sured indirectly by the BOLD signal, and the temporal regularity in the stimulus. These analyses assess the significance of the comparison within each of the brainstem structures, with correction for the volume of those structures. The contrast between the regular-interval-sound and noise matched in intensity and passband was significant in both cochlear nuclei at the  $p < 0.05$  level, while the same contrast was significant in both inferior colliculi at the  $p < 0.005$  level. The study therefore represents a demonstration of an increase in the local mean firing rate as a function of the stimulus regularity as early as the cochlear nucleus, with a more significant relationship in the inferior colliculus. What does this mean? In the cochlear nucleus there are probably two possibilities. One is that there may be a subpopulation of cells that increase in mean firing rate in response to particular temporal regularities corresponding to particular pitches. However, neurophysiological studies in the guinea pig (Winter, Chapter 4) have not demonstrated such a selective response; the responses to temporal regularity in onset choppers are selective but the selectivity is demonstrated in the temporal firing pattern rather than the mean rate. A second possibility in the cochlear nucleus is that the mean firing rate in a larger population of neurons increases as a function of *synchronization* of the local networks of neurons due to the regularity of the stimulus. Relevant modeling studies (Chawla et al. 1999) were motivated by a need to study cortical processing, but used networks of excitatory and inhibitory neurons with conventional Hodgkin–Huxley neural dynamics and a pattern of interconnections that could plausibly be applied to brainstem nuclei. The studies demonstrated tight coupling of mean activity levels and synchronization that was not sensitive to large changes in the model parameters. On the basis of the absence of a candidate cell in the cochlear nucleus with a tuned rate response to regular interval sound, the synchronization mechanism for increasing the local BOLD response would seem more plausible. In the case of either possible mechanism in the cochlear nucleus, the more significant relationship in the inferior colliculus points to a stabilized neural representation at that level that is based on a local rate code. Such a representation is predicted by physiological models such as that of Langner (1992) and the psychophysical auditory image model (AIM) of Patterson et al. (1995). The Langner model is specific about such a representation in the inferior colliculus while the original model of Patterson model was not as anatomically constrained.

Although this functional MRI work can demonstrate the vertical level at which temporal regularity is converted to a rate code, the technique does not have the anatomical precision to demonstrate the systematic mapping of temporal structure in the inferior colliculus suggested in the Langner model. The study used typical spatial smoothing with a filter with full width at half maximum of 5 mm, which is probably an order of magnitude too coarse to test hypotheses about periodicity maps, at least in the brainstem.

This discussion of brainstem processing relevant to pitch perception has concentrated on temporal processing. This is not to dismiss the relevance of spectral encoding to the perception of pitch, especially at low frequencies, where the



presence of the resolved lower harmonics increases pitch salience. In the brainstem, demonstration of tonotopy suffers from the same problem as demonstration of the mapping of regularity: the lack of anatomical resolution. Melcher and colleagues at Massachusetts General Hospital have seen trends consistent with tonotopic organization in the inferior colliculus (Melcher, unpublished observation) but no published study to date has demonstrated any systematic mapping. The brainstem mapping of tonotopy in mammalian neurophysiological studies by Langner and others is one form of indirect argument for its existence in humans. A much stronger argument is the tonotopy that has been demonstrated in the human cortex, described in Section 2.3. This could not occur without a preservation of tonotopic mapping in the human brainstem.

### *2.3 Processing of Stimulus Properties Relevant to Pitch in the Auditory Cortex*

#### *2.3.1 Spectral Representation*

In terms of the spectral structure of sound relevant to pitch, a number of studies have addressed the question of tonotopic mapping in the auditory cortex. In humans, the auditory cortex is located in the superior temporal plane, the superior surface of the temporal lobe within the Sylvian fissure. The superior temporal plane is conveniently shown in a tilted axial section like Figure 5.2. The primary auditory cortex is located in the region of Heschl's gyrus (HG). HG is a gyrus running laterally and anteriorly in the superior temporal plane. There is a degree of macroscopic variability in that there might be one, two, or even three HGs, and the number can differ on either side. Detailed human anatomical studies confirm that primary auditory cortex, defined on the basis of microscopic structure (cytoarchitectonics), is related to the medial part of HG (or the most anterior HG if there are more than one). However, there is considerable anatomical variation of the cytoarchitectonic area with respect to the macroscopic boundaries, whereby as little as 30% or as much as 80% of the gyrus can be taken up by primary cortex (Morosan et al. 2001). It is important to bear this in mind when considering imaging studies of the cortex. When these are able to look at activation in individual subjects, lack of consistency between subjects may partly reflect the fact that such consistency can be defined only with respect to macroscopic landmarks, rather than with respect to the functional auditory areas that are not accurately delineated by such landmarks.

Lauter et al. (1985) carried out the first study to examine tonotopic organization in the human cortex using PET. A comparison was made in a group study between activation based on pure tones at 500 Hz and 4 kHz. Both tones produced activation in the superior temporal plane, with more lateral activation being produced by the lower-frequency tone. In a later PET study, Lockwood and colleagues (Lockwood et al. 1999) found two foci of activation in the left

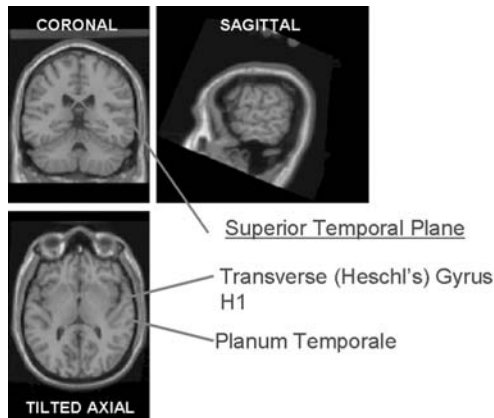


FIGURE 5.2. Anatomy of human auditory areas. Tilted axial section at the level of the superior temporal plane allows definition of the primary and secondary auditory areas. Also shown in this figure are coronal and sagittal sections at the level of the auditory cortex. The primary auditory cortex corresponds to the medial part of Heschl's gyrus (shaded red), but note that there is no exact correspondence between the cytoarchitectonically defined areas and the macroscopic boundaries (see text). (See color insert.)

hemisphere (in medial and lateral HG) for a 4-kHz pure tone presented to the right ear at 90 dB hearing level. For a 500-Hz tone, a single focus of activation was demonstrated in the lateral part of HG, at the same point as the lateral focus for the 4kHz tone. The distinct patterns produced by the two tones provide evidence for a tonotopic mapping in the superior temporal plane. However, the precise pattern of mapping is difficult to demonstrate in PET experiments based on group data where spatial smoothing rarely exceeds a filter width at half-maximum of 10 mm.

A number of studies have employed fMRI to investigate tonotopic mapping in the cortex (e.g., Wessinger et al. 1997; Talavage et al. 2000) where increased spatial resolution and the ability to carry out individual analyses are a particular advantage. A disadvantage of fMRI is the scanner noise; both the studies of Wessinger et al. and Talavage et al. used "epoch mode" designs where there is continuous acquisition of data (and therefore scanner noise) during presentation of the stimuli of interest. In the Wessinger study harmonic tones with most spectral energy at 55 Hz or 880 Hz were presented diotically to subjects. A consistent pattern of activation was demonstrated in the left hemispheres of subjects in whom the activation due to the high-frequency tone was more medial in the superior temporal plane than the activation due to the low frequency tone, but the same consistency was not observed in the right hemispheres of the subjects. Talavage et al. used a variety of stimuli where the spectral distribution could be varied (pure tones at 650 Hz and 2.5 kHz, 10-Hz amplitude-modulated tones with same carrier frequencies, and broadband stimuli [AM noise and mu-

sic] that were low- or high-pass filtered). For each type of stimulus type there was a low- and a high-frequency condition. Areas were defined where there was a greater response to either the high or the low frequency. A low-frequency area was identified in HG, while high-frequency areas on HG were identified both lateral and medial to it. Talavage et al. proposed that the organization of responses along HG is consistent with having two adjacent tonotopic maps with mirror reversal between them at the low-frequency point. Further, they proposed that the medial and lateral areas correspond, respectively, to areas A1 and R in the macaque (Merzenich and Brugge 1973; Kaas and Hackett 2000). A similar mirror reversal of tonotopy at low frequency is seen between A1 and R in the macaque studies.

Additional areas in the Talavage et al. study may correspond to anterior, posterior, and lateral areas identified in human anatomical studies of auditory cortex (Rivier and Clarke 1997). Although the degree of homology between human and macaque is an open question, the human imaging studies strongly support the existence of distinct tonotopic mappings in different areas within the superior temporal plane. Such mappings afford a mechanism for the representation of spectral sound properties relevant to pitch perception.

### 2.3.2 Temporal Representation

An early study with regular-interval sounds used PET to identify cortical areas showing a relationship between the stimulus regularity and the local activity as measured by regional cerebral blood flow (Griffiths et al. 1998). This work showed activation in HG as a function of the temporal regularity of the stimulus. On the basis of such a PET group study it is not possible to say whether the peak activation as a function of stimulus regularity was in primary cortex in medial HG or secondary cortex in lateral HG, and PET does not allow examination of individual data. A recent fMRI group analysis of the cortical BOLD response to temporal regularity (Patterson et al. 2002) has demonstrated bilateral peak activation as a function of temporal regularity in secondary auditory cortex in lateral HG. This was demonstrated in both group analyses (Fig. 5.3) and in eight of the nine individual analyses (Fig. 5.4). Notice the remarkably consistent individual findings in Figure 5.4, where the red activation corresponding to the contrast between regular-interval sound and a matched noise is located in lateral HG.

## 2.4 Evidence for a “Pitch Center” in Secondary Auditory Cortex

The cortical data showing activation as a function of stimulus regularity beg the question: What is represented in lateral HG when the temporal regularity of the stimulus is varied? As argued above, it is not possible to make a definitive statement on the basis of the data. The activation corresponds to a population

rate code that might be related to temporal regularity in the stimulus or to the percept of pitch. An argument in favor of the latter idea, albeit weak and indirect, is: Why should such stimulus representations exist at such an advanced point in the cortical auditory system, when they first occur in the brainstem? Furthermore, direct evidence in favor of a “pitch center” comes from another experiment in which pitch salience is varied in a different manner. In an fMRI

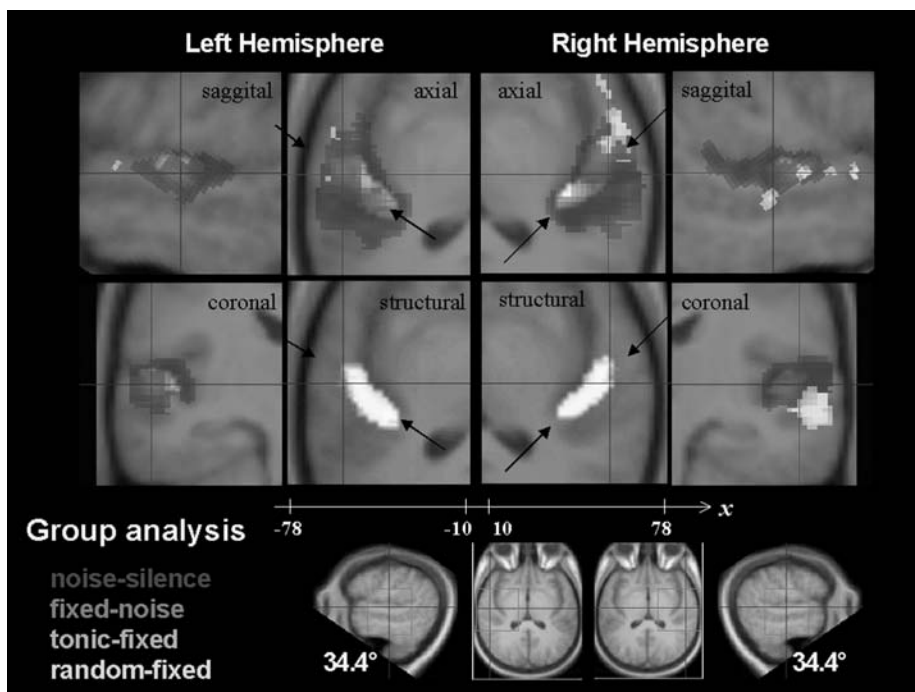


FIGURE 5.3. fMRI activation for contrasts between noise stimuli with different temporal regularity and pitch strength, and with different pitch patterns. Group data for nine subjects are shown. The contrasts are rendered onto the average structural image of the group. Blue: activation in response to noise bursts (versus silence); red: differential activation in response to notes with fixed pitch (versus noise bursts); green: differential activation in response to tonic melodies (versus fixed pitch); cyan: differential activation in response to random melodies. The white area shows the mean position of Heschl's gyrus for the group. The arrows show the midline of Heschl's gyrus separately in each hemisphere. The position and orientation of the sections are illustrated in the bottom panels of the figure. The “axial” section is tilted by 0.6 radians (or 34.4°) relative to the horizontal plane to show the entire surface of the temporal lobe in one plane. The other sections are sagittal and coronal with respect to the surface of the temporal lobe. The sagittal sections show front to the left for the left hemisphere and front to the right for the right hemisphere, that is, they are being viewed from outside the brain volume. (See color insert.) Reproduced from Patterson et al. (2002) with permission from Elsevier.

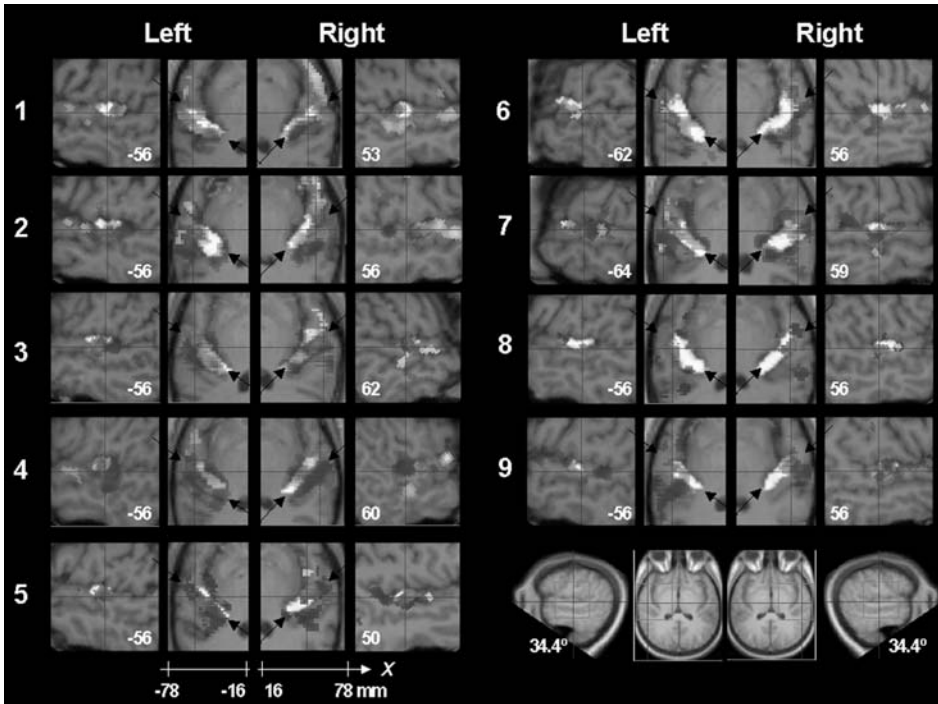


FIGURE 5.4. fMRI activation for the same contrasts as Figure 5.3, this time shown for nine individual listeners rendered on sections of their individual structural images. The orientation of the “axial” sections is the same as in Figure 5.3. The plane of each sagittal section is given in mm in Talairach space (Talairach and Tournoux 1988) in each of the respective panels. The position of each individual’s Heschl’s gyrus is highlighted in white in each case. The pairs of *black arrows* in the axial sections of each row show the position of the average Heschl’s gyrus; that is, they are the same arrows as in the central panels of the upper row of Figure 5.3. Blue: noise activation (versus silence); red: fixed-pitch activation (versus noise); green: combined differential activation to tonic and random melodies (versus fixed pitch). (See color insert.) Reproduced from Patterson et al. (2002) with permission from Elsevier.

experiment, Penagos et al. (2003) showed that activation in lateral HG resulting from harmonic stimuli within a particular passband increased when the harmonics were resolved, and the pitch salience was greater, than when the harmonics were unresolved and the pitch salience was weaker. Stimulus regularity was the same in both cases, as both were exactly periodic stimuli.

An argument about a human homolog of area R in the macaque in lateral HG has already been made on the basis of the tonotopic data of Talavage et al. (2000). Patterson et al. (2002) also speculate about the existence of such a homolog and suggest that a neural correlate of the percept of pitch may exist in this center.

## *2.5 Distributed Temporal Lobe Mechanisms for the Processing of Pitch Patterns*

The synthesis so far has developed the idea that processing of the pitch of individual sounds depends on hierarchical mechanisms. Stabilized representations of stimulus properties relevant to pitch are constructed in the brainstem and used to form a cortical representation corresponding to the percept of pitch. In the real acoustic world we do not listen to single pitches but to patterns of pitch such as the fundamental frequency (F0) contour of speech, relevant to stress and prosody, and the melody of music. These patterns exist within a temporal window of seconds or tens of seconds rather than the temporal window of milliseconds relevant to the processing of temporal regularity within individual sounds. This section considers how the brain represents these patterns.

In the work using regular-interval sounds with associated pitch described above (Patterson et al. 2002), the individual sounds containing regularity were presented in sequences with both fixed pitch and as two types of pitch pattern; a random variation of pitch and as a tonal melody. Comparison between the sequences containing the pitch variation produced a pattern of cortical activation that was distinct from the pattern produced by the fixed pitch. A striking negative finding here was the absence of any significant activation in the brainstem nuclei. The processing of pitch patterns within this long temporal window involves the cortex, and the network of areas activated is distinct from the region of the primary auditory cortex. Bilateral activation was demonstrated in the posterior superior temporal lobe (in the region of the planum temporale and adjacent superior temporal gyrus) and in the anterior superior temporal lobe in the planum polare (Fig. 5.3 and 5.4). This is the first level of pitch processing at which asymmetries emerge, with a greater extent and significance for the right-sided activations. The other feature to emerge at this level of pitch processing is a much greater variation between individuals. Compare in Figure 5.4 the highly conserved activation in red due to fixed pitch with the much more variable activation due to the imposition of a pitch pattern. The four areas demonstrated in the group analysis in this experiment were similar to four areas demonstrated in a previous interaction analysis in a PET study that was designed to demonstrate areas interested in “long-term” temporal structure. However, the previous study did not demonstrate the asymmetry that was apparent in the recent fMRI study. The right lateralization during the processing of pitch pattern would be consistent with studies of melody perception and imagery (Zatorre et al. 1994, 1996).

A surprising aspect of the fMRI cortical study was the absence of any marked differences between the random pitch and melody conditions. Although the melodies were novel and would not have semantic associations, they were tonal melodies with a contour or long-term structure not present in the random pitch condition. Cognitive neuropsychology suggests that the processing of the “local” structure of pitch sequences has a different basis from the processing of contour (Liegeois-Chauvel et al. 1998). Of possible relevance here is the ab-

sence of any task in the fMRI experiment that was primarily concerned with pitch perception. A number of experiments where comparison tasks for pitch sequences are employed (e.g., Zatorre et al. 1994; Griffiths et al. 1999) have shown frontal activation not seen in the recent experiment. It is conceivable that differences between the cortical processing of different types of pitch pattern may only emerge when the brain has to make use of them.

### 3. Electromagnetic Techniques: EEG and MEG

#### *3.1 Background: The Basis for (and Limitations of) Electromagnetic Techniques*

Electromagnetic techniques (EEG and MEG) allow the recording of brain responses to sound stimuli with millisecond accuracy. Spatial precision is becoming increasingly high in MEG work (see, e.g., Fig. 5.5), but spatial localization requires a number of assumptions in both EEG and MEG. In the case of EEG, sources can be modeled using a Laplacian transformation, and in the case of MEG sources can be modeled by the fitting one or more equivalent current dipoles as generators of the magnetic field measured outside the head. There is general agreement that the magnetic field measured using MEG is generated by currents in the dendrites of cortical neurons. Based on the arguments in the previous section, this means that both fMRI and MEG reflect similar neuronal mechanisms, but over very different time scales. MEG is sensitive mainly to sources oriented tangentially with respect to the surface of the skull. This is often the case for sources in the auditory areas within the superior temporal plane (see, e.g., Fig. 5.5) and robust responses to auditory stimuli can be measured. Indeed, in a number of laboratories one of the auditory responses, the N1m, is used as a form of quality control.

MEG suffers from the problem of nonunique solutions to the inverse problem. This means that, for a given pattern of magnetic activity around the skull there are an infinite number of possible combinations of equivalent current dipoles within the skull that might produce such a pattern. There are a number of approaches to dealing with this problem. One is to constrain the set of possible equivalent current dipoles by prior assumptions. These might be uncontroversial, such as the assumption that the dipoles are within the brain, or more constrained, such as the assumption that the dipole must be in the cortex. Some studies constrain the dipoles to be in a particular part of the brain, which carries a risk of making tautological arguments about the origin of the activity demonstrated. Another approach that may be reasonable for early auditory areas is to assume a single equivalent dipole accounting for the activity in each hemisphere. This approach might be based on a form of least-squares fitting or maximum likelihood estimation, and can be influenced by the seeding of the dipole search (i.e., where the algorithm starts to look).

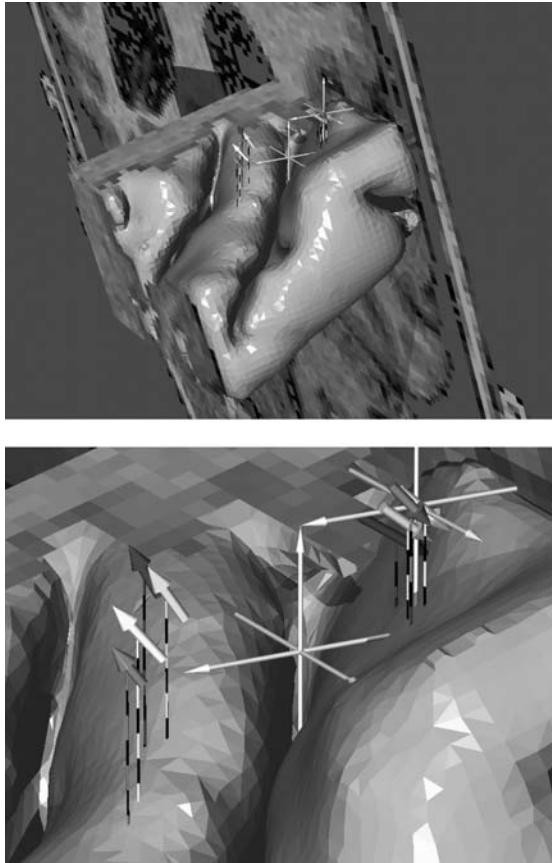


FIGURE 5.5. Three-dimensional reconstruction of part of the left superior temporal plane in a detailed single-subject study. The middle ridge corresponds to Heschl's gyrus and the area behind (on the *right* in the figure) to the planum temporale. The *lower part* of the figure is a magnified version of the *upper*. The *arrows* correspond to the equivalent current dipoles at different frequencies (red, yellow, green, and blue correspond to 250, 500, 1000, and 2000 Hz, respectively). The orientation of the *arrows* is shown above the cortical surface and is connected to the point on the cortical surface where the dipole is located by a *vertical line*. The dipoles on the planum temporale on the right correspond to the N1m response with a latency of about 100 ms. The dipoles above Heschl's gyrus on the left correspond to the P2m with a latency of 150 to 200 ms. Tonotopic mapping is demonstrated with millimeter precision where high-frequency responses are represented more medially in the planum temporale for the N1m and in Heschl's gyrus for the P2m. (See color insert.) Reproduced with permission from Figure 6a and c in Lütkenhöner and Steinsträter (1998). © S. Karger AG, Basel.



### *3.2 Responses to Clicks in the Ascending Auditory Pathway and Superior Temporal Plane*

The mapping of electrical responses to single clicks in the ascending auditory pathway and cortex is not directly relevant to the processing of spectral or time-domain properties relevant to pitch perception. However, the responses recorded at different latencies provide a basis for comparison with studies using more sophisticated stimuli associated with pitch. In the ascending auditory pathway EEG responses to clicks, the brainstem auditory evoked potentials, are well described, and arise from structures between the auditory nerve (waves I and II) and the lateral lemniscus/inferior colliculus (wave V). These responses all occur at latencies less than 10 ms. MEG does not allow reliable recording of responses from deep structures located in the brainstem.

In the superior temporal plane EEG studies using depth electrodes in subjects with epilepsy provide direct evidence that an electrical response from auditory cortex in medial HG can occur with latencies as short as 20 ms (Liegeois-Chauvel et al. 1991; Howard et al. 2000). Scalp recorded EEG studies and MEG studies of click responses can show a response at a similar latency called the Na (EEG) or Nam (MEG), which in the case of the EEG Na response is the first vertex negative wave of what are called the middle latency responses. The response can be fickle when recorded using MEG, although Lutkenhoner et al. (2003a) suggest that consistent MEG Nam responses can be achieved by using the first temporal derivative of the signal. The other components of the middle latency responses (Pa, Nb, Pb for EEG; Pam, Nbm, Pbm for MEG) have latencies of up to 70 ms. Pam arises from the medial part of HG, and Nbm and Pbm from the lateral part of HG (Yvert et al. 2001). The N1 or N100 (N1m or N100m for MEG) is a robust vertex-negative response in the EEG with a latency of approximately 100 ms. MEG mapping demonstrates an origin for the N1m in the planum temporale for this component. The P2 or P2m is a later response occurring at a latency of about 150 to 200 ms. Source localization studies show that the P2m wave source is anterior to N1, with a probable origin in HG. The response to click stimulation mapped with MEG can therefore be seen to change with increasing latency from medial HG, to lateral HG, to the planum temporale behind and then anteriorly back on to HG.

### *3.3 Tonotopy in the Superior Temporal Plane*

Studies based on tone stimulation rather than clicks allow an examination of tonotopy in the superior temporal plane. A number of workers have addressed this issue using MEG and this section illustrates the approach with certain examples. Pantev et al. (1995) showed a dependence of both Pam and N1m on frequency using tones of 500, 1000, and 4000 Hz. The tonotopic mapping was mirrored for the Pam and N1m responses. The Pam response in HG became more lateral (superficial) with increasing frequency while the N1m response in PT became more medial (deeper) with increasing frequency. In a detailed

single-subject study Lütkenhöner and Steinsträter (1998) investigated both N1m and the later P2m responses (Fig. 5.5). A similar mapping of the N1m was demonstrated as in previous studies, with generators of responses to higher frequencies occurring more medially in PT. A tonotopic arrangement of dipoles was also observed in the mapping of the P2m in HG. However, it was argued that these responses might not be adequately described by a single generator. This reservation was corroborated by recent data from 19 hemispheres (Lütkenhöner et al. 2003b), suggesting that the investigation of the tonotopy of N1m is problematic.

Like the fMRI data, the MEG data support the existence of multiple areas within the superior temporal plane that are tonotopically organized. The earliest response for which tonotopy has been demonstrated, Pa, is not the earliest cortical response to sound, which is Na, although Pa does arise from the medial part of HG where cytoarchitecturally defined primary cortex is located (Morosan et al. 2001). The MEG data suggest that the tonotopic generators of early “more primary” responses (Pam) and later responses (P2m) both map to HG, and suggest the need for caution when interpreting the tonotopic data based on fMRI BOLD responses acquired over much longer time windows.

### *3.4 Responses to Temporal Regularity in the Superior Temporal Plane*

In an early MEG study Pantev et al. (1989) mapped the N1m response in the superior temporal plane as a function of the pitch of a missing fundamental stimulus. The work suggested that the N1m due to both pure tones and to missing fundamental stimuli with matched pitch showed a similar mapping. In both cases, the generators for stimuli associated with a higher pitch were estimated to be deeper. However, the original result was not replicated in a recent study (Lütkenhöner 2003), possibly reflecting a problem with interpretations based on the use of a single-dipole model for the N1m. The original work was interpreted in terms of the mapping of a neural correlate of the conscious perception of pitch, rather than frequency. This would be a parsimonious explanation for the data, although the pitch of the missing fundamental corresponds to the periodicity of the stimulus and the mapping could also be explained on the basis of time-domain stimulus properties. In another study using the missing fundamental stimulus, where the spectral passband and F0 were manipulated independently, Langner and colleagues (Langner et al. 1997) argued for a distinct orthogonal mapping of temporal and spectral sound characteristics in the superior temporal plane.

Gutschalk et al. (2002) compared the responses to very rapid click trains that were either periodic and associated with pitch or random. A sustained magnetic field was demonstrated to prolonged click trains with a generator in lateral HG was that was different for the regular and irregular sounds at click rates above 40 Hz (above 40 Hz the regular clicks have a strong associated pitch). The decrease of the differential response to pitch with decreasing repetition rate

occurs at rates where the pitch salience decreases (Krumbholtz et al. 2000). This represents further circumstantial evidence that a neural correlate of pitch perception exists in lateral HG, in accord with the suggestion from the fMRI study of regular interval sounds (Patterson et al. 2002) and another recent MEG study (Krumbholtz et al. 2003).

### *3.5 Analysis of Differences in Pitch Responses Between Subjects*

Patel and Balaban (2001) used independent manipulation of spectral passband and F0 in harmonic sounds to create stimuli where the spectral passband and F0 could go in opposite directions. The stimuli were heard by some subjects as rising in pitch and by others as decreasing in pitch, depending on whether they heard the missing fundamental or not. These stimuli allow a demonstration of the perceptual representation of pitch. A steady-state technique was employed in which amplitude modulation of the harmonic sounds at 41.5 Hz was used as a marker of the neural response to the signal. The study was not primarily designed to map the origin of the responses in the same way as the studies of transient MEG responses described above, but allowed the comparison of responses arising from the two hemispheres. For all subjects, responses in the magnitude and phase spectra at 41.5 Hz could be demonstrated in both hemispheres that changed in a different sense depending on the direction of F0. The fact that these responses occurred in subjects regardless of whether they heard the missing fundamental or not suggests a sensory mapping, the most likely basis being one corresponding to the periodicity of the signal. Differences between the subjects who perceived the missing fundamental and those who did not were shown in the phase spectra in the right hemisphere only; phase responses in the right hemisphere that changed according to the direction of F0 were significantly less common in subjects who did not perceive the complex pitch. This is interesting in view of the cognitive neuropsychological literature suggesting that right temporal lobe lesions have a particular effect on the perception of the pitch of the missing fundamental (Zatorre 1988).

### *3.6 Responses to Pitch Patterns*

The discussion so far about the use of electromagnetic imaging techniques has focused on what the techniques can reveal about the processing of stimulus properties relevant to pitch and pitch salience. These techniques also allow an examination of the processing of *patterns* of pitch, at a higher level of complexity. Hemodynamic techniques have the advantage of looking at a long temporal window several seconds long during which time there might be changes that correspond to the contour or long-term structure of a pitch sequence. A disadvantage of the techniques, however, is that it is very difficult to design experiments where the “local” structure of pitch sequences (absolute pitch values and intervals) and the “global” structure or contour can be manipulated inde-

pently. I use local and global here in the same sense as Dowling and Harwood (1985), who developed psychophysical tests of local and global processing based on the comparison of pitch sequences containing local changes in pitch (alteration in one pitch without changing the overall pattern of ups and downs or contour) and global changes in pitch (where the contour changes). Schiavetto et al. (1999) carried out an interesting EEG study in which they altered one pitch in a sequence in either a contour-preserved (local) or contour-violated (global) condition. They demonstrated an N2 response at 200 ms to the global change (as assessed by the difference between the sequences with an altered pitch at one fixed point and the standards) that peaked in frontocentral regions and also a frontal P3b response at 300 ms. The local response only produced a P3b response. These data suggest widely distributed brain processes including frontal processing for the analysis of pitch pattern. Such widely distributed processing is also suggested by studies using hemodynamic techniques and melodies (Zatorre et al. 1994; Patterson et al. 2002). However, the electromagnetic studies can go further in allowing the fractionation of local and global processing. The Schiavetto et al. data can be interpreted in terms of the Dowling and Harwood (1985) model of pitch perception, where there is a primary processing of global structure to produce a cognitive structure before local details are “hung” onto it.

Patel and Balaban (2000) used their MEG method to demonstrate neural responses that “track” the pitch contour of sound sequences. They produced sequences of tones with fixed modulation rate of 41.5 Hz and varying pitch determined by the carrier. The modulation was used as a “marker” for the neural response to the signal; the response to successive notes was assessed based on the amplitude and phase spectrum at 41.5 Hz. MEG responses were demonstrated where the phase response “followed” the pitch sequence over time, and where the tracking became more accurate as the sequence became less random. Coherence between responses in different brain regions were also demonstrated; this long-term coherence between areas was greatest when pitch sequences were used that had a similar combination of contour and local variation to musical pitch patterns.

#### 4. Conclusion

Considered as a whole, the hemodynamic studies and electromagnetic studies are consistent with a hierarchy of pitch processing in humans in which (1) spectral and temporal features of sounds relevant to pitch are encoded in the brainstem, (2) a neural correlate of the conscious perception of pitch exists in areas of auditory cortex distinct from the primary auditory cortex, and (3) longer time scale patterns of pitch are processed in networks including areas in the temporal lobes (distinct from the primary and secondary areas) and in the frontal lobes.

A number of questions remain regarding the human processing of pitch. The

understanding of both the hemodynamic techniques and electromagnetic techniques is predicated on the anatomy of the superior temporal plane and our knowledge of this is evolving rapidly at the moment. There is no reliable way to demonstrate the microscopic areas in humans reliably *in vivo*, to compare these with the functional areas derived from functional imaging techniques. It is possible that emerging techniques such as diffusion-tensor MRI imaging may help with such distinctions, as well as with the other critical gap in our understanding; the pattern of connectivity of the different areas. The electrical techniques already go some way to suggesting the pattern of sequential activation in cortical areas. The hope is that the anatomy and functional imaging techniques continue to produce a convergent picture of the exact mapping of stimulus features and neural correlates of perception. This will allow us to constrain hypotheses about pitch perception based on psychophysics and also compare humans with animals in situations where we have a much more detailed knowledge of the neurophysiology.

The largest gap in our knowledge in this area relates to the use of pitch in the sort of sound patterns we experience in the real acoustic world. This chapter has touched on sound-sequence analysis, where it is already clear that no single imaging technique is adequate in its own right to answer questions about the neural substrates for the high-level processing of patterns of pitch. Pitch processing is highly relevant to the analysis of sound objects and sound streams and it is perhaps in this area that the greatest amount of work remains to be done.

*Acknowledgments.* My work is supported by the Wellcome Trust (UK). All the pitch studies in which I have been involved were carried out at the Wellcome Department of Imaging Neuroscience, London, UK. The work on regular interval sound was carried out with other members of the Centre for the Neural Basis of Hearing, Cambridge (Roy Patterson and Stefan Uppenkamp) and with Ingrid Johnsrude, Cambridge University.

## References

- Chawla D, Lumer ED, Friston KJ (1999) The relationship between synchronization among neuronal populations and their mean activity levels. *Neur Comput* 11:1389–411.
- Dowling WJ, Harwood DL (1985) *Music and Cognition*. London: Academic Press.
- Frith C, Perry R, Lumer E (1999) The neural correlates of conscious experience: an experimental framework. *Trends Cogn Sci* 3:105–114.
- Griffiths TD, Buechel C, Frackowiak RSJ, Patterson RH (1998) Analysis of temporal structure in sound by the human brain. *Nat Neurosci* 1:421–427.
- Griffiths TD, Johnsrude I, Dean JL, Green GGR (1999) A common neural substrate for the analysis of pitch and duration pattern in segmented sound? *NeuroReport* 18:3825–3830.

- Griffiths TD, Uppenkamp S, Johnsrude I, Josephs O, Patterson RD (2001) Encoding of the temporal regularity of sound in the human brainstem. *Nat Neurosci* 4:633–637.
- Guimares AR, Melcher JR, Talavage TM, Baker JR, Ledden P, Rosen BR, Kiang NYS, Fullerton BC, Weisskoff RM (1998) Imaging subcortical activity in humans. *Hum Brain Map* 6:33–41.
- Gutschalk A, Patterson RD, Rupp A, Uppenkamp S, Scherg M (2002) Sustained magnetic fields reveal separate sites for sound level and temporal regularity in human auditory cortex. *NeuroImage* 15:207–216.
- Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW (1999) “Sparse” temporal sampling in auditory fMRI. *Hum Brain Map* 7:213–223.
- Howard MA, Volkov IO, Mirsky R, Garell PC, Noh MD, Granner M, Damasio H, Stein-schneider M, Reale RA, Hind JE, Brugge JF (2000) Auditory cortex on the human posterior superior temporal gyrus. *J Comp Neurol* 416:79–92.
- Kaas JH, Hackett TA (2000) Subdivision of auditory cortex and processing streams in primates. *Proc Natl Acad Sci USA* 97:11793–11799.
- Krumbholz K, Patterson RD, Pressnitzer D (2000) The lower limit of pitch as determined by rate discrimination. *J Acoust Soc Am* 108:1170–1180.
- Krumbholz K, Patterson R.D, Seither-Preisler A, Lammertmann C, Lütkenhöner B. (2003) Neuromagnetic evidence for a pitch processing centre in Heschl’s gyrus. *Cereb Cortex* 13:765–772.
- Langner G (1992) Periodicity encoding in the auditory system. *Hear Res* 60:115–142.
- Langner G, Sams M, Heil P, Schulze H (1997) Frequency and periodicity are represented in orthogonal maps in the human auditory cortex: evidence from magnetoencephalography. *J Comp Physiol A* 181:665–676.
- Lauter JL, Herscovitch P, Formby C, Raichle ME (1985) Tonotopic organization in the human auditory cortex revealed by positron emission tomography. *Hear Res* 20:199–205.
- Liegeois-Chauvel C, Musolino A, Chauvel P (1991) Localization of the primary auditory area in man. *Brain* 114:139–153.
- Liegeois-Chauvel C, Peretz I, Babai M, Laguittin V, Chauvel P (1998) Contribution of different cortical areas in the temporal lobes to music processing. *Brain* 121:1853–1867.
- Lockwood AH, Salvi RJ, Coad ML, Arnold SA, Wack DS, Murphy BW, Burkard RF. (1999) The functional anatomy of the normal human auditory system: responses to 0.5 and 4.0kHz tones at varied intensities. *Cereb Cortex* 9:65–76.
- Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A (2001) Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412:150–157.
- Lumer ED, Friston KJ, Rees G (1998) Neural correlates of perceptual rivalry in the human brain. *Science* 280:1930–1934.
- Lütkenhöner B (2003) Single-dipole analyses of the N100m are not suitable for characterizing the cortical representation of pitch. *Audiol NeuroOtol* 8:222–233.
- Lütkenhöner B, Steinsträter O (1998) High-precision neuromagnetic study of the functional organization of the human auditory cortex. *Audiol NeuroOtol* 3:191–213.
- Lütkenhöner B, Krumbholz K, Lammertmann C, Seither-Preisler A, Steinsträter O, Patterson RD (2003a) Localization of primary auditory cortex in humans by magnetoencephalography. *NeuroImage* 18:58–66.
- Lütkenhöner B, Krumbholz K, Seither-Preisler A (2003b) Studies of tonotopy based

- on wave N100 of the auditory evoked field are problematic. *NeuroImage* 19:935–949.
- Merzenich MM, Brugge JF (1973) Representation of the cochlear partition on the superior temporal plane of the macaque monkey. *J Neurophysiol* 24:193–202.
- 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, Lütkenhöner B, Lehnertz K (1989) Tonotopic organisation of the auditory cortex: pitch versus frequency representation. *Science* 242:486–488.
- 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. *EEG Clin Neurophysiol* 94:26–40.
- Parri R, Crunelli V (2003) An astrocyte bridge from synapse to blood flow. *Nat Neurosci* 6:5–6.
- Patel AD, Balaban E (2000) Temporal patterns of human cortical activity reflect tone sequence structure. *Nature* 404:80–84.
- Patel AD, Balaban E (2001) Human pitch perception is reflected in the timing of stimulus-related cortical activity. *Nat Neurosci* 4:839–844.
- Patterson RD, Allerhand MH, Giguere C (1995) Time-domain modeling of peripheral auditory processing: a modular architecture and a software platform. *J Acoust Soc Am* 98:1890–1894.
- Patterson RD, Uppenkamp S, Johnsrude I, Griffiths TD (2002) The processing of temporal pitch and melody information in auditory cortex. *Neuron* 36:767–776.
- Penagos H, Melcher JR, Oxenham AJ (2004) A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *J Neurosci* 24:6810–6815.
- Ravicz ME, Melcher JR, Kiang NY (2000) Acoustic noise during functional magnetic resonance imaging. *J Acoust Soc Am* 108:1683–1696.
- 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.
- Schiavetto A, Cortese F, Alain C (1999) Global and local processing of musical sequences: an event-related brain potential study. *NeuroReport* 10:2467–2472.
- Seifritz E, Esposito F, Hennel F, Mustovic H, Neuhoff JG, Bilecen D, Tedeschi G, Scheffler K, Di Salle F (2002) Spatiotemporal pattern of neural processing in the human auditory cortex. *Science* 297:1706–1708.
- Talairach P, Tournoux J (1988) *A Stereotactic Coplanar Atlas of the Human Brain*. Stuttgart: Thieme.
- Talavage TM, Ledden PJ, Benson RR, Rosen BR, Melcher JR (2000) Frequency-dependent responses exhibited by multiple regions in human auditory cortex. *Hear Res* 150:225–244.
- Wessinger CM, Buonocore MH, Kussmaul CL, Mangun GR (1997) Tonotopy in human auditory cortex examined with functional magnetic resonance imaging. *Human Brain Map* 5:18–25.
- Yost WA, Patterson R, Sheft S (1996) A time domain description for the pitch strength of iterated rippled noise. *J Acoust Soc Am* 99:1066–1078.
- Yvert B, Crouzeix A, Bertrand O, Seither-Preisler A, Pantev C (2001) Multiple supra-

- temporal sources of magnetic and electric auditory evoked middle latency components in humans. *Cereb Cortex* 11:411–423.
- Zatorre R (1988) Pitch perception of complex tones and human cerebral lobe function. *J Acoust Soc Am* 84:566–572.
- Zatorre RJ, Evans AC, Meyer E (1994) Neural mechanisms underlying melodic perception and memory for pitch. *J Neurosci* 14:1908–1919.
- Zatorre RJ, Halpern AR, Perry DW, Meyer E, Evans AC (1996) Hearing in the mind's ear- a PET investigation of musical imagery and perception. *J Cogn Neurosci* 8:29–46.



# 6

## Pitch Perception Models

ALAIN DE CHEVEIGNÉ

### 1. Introduction

This chapter discusses models of pitch, old and recent. The aim is to chart their common points—many are variations on a theme—and differences, and build a catalog of ideas for use in understanding pitch perception. The busy reader might read just the next section, a crash course in pitch theory that explains why some obvious ideas do not work and what are currently the best answers. The brave reader will read on as we delve more deeply into the origin of concepts and the intricate and ingenious ideas behind the models and metaphors that we use to make progress in understanding pitch.

### 2. Pitch Theory in a Nutshell

Pitch-evoking stimuli usually are periodic, and the pitch usually is related to the period. Accordingly, a pitch perception mechanism must estimate the period  $T$  (or its inverse, the fundamental frequency  $F_0$ ) of the stimulus. There are two approaches to do so. One involves the *spectrum* and the other the *waveform*. The two are illustrated with examples of stimuli that evoke pitch, such as pure and complex tones.

#### 2.1 *Spectrum*

The spectral approach is based on Fourier analysis. The spectrum of a pure tone is illustrated in Figure 6.1A. An algorithm to measure its period (inverse of its frequency) is to look for the spectral peak and use its position as a cue to pitch. This works for a pure tone, but consider now the sound illustrated in Figure 6.1B, which evokes the same pitch. There are several peaks in the spectrum, but the previous algorithm was designed to expect only one. A reasonable modification is to take the largest peak, but consider now the sound illustrated in Figure 6.1C. The largest spectral peak is at a higher harmonic, yet the pitch

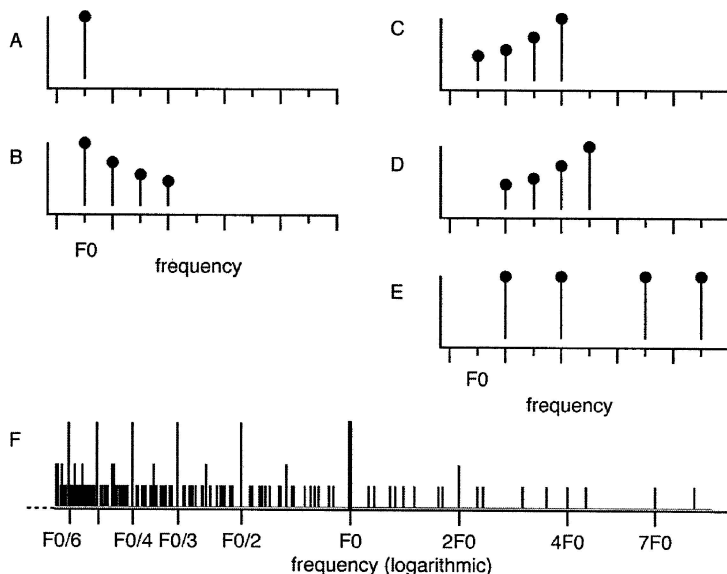


FIGURE 6.1. Spectral approach. (A) to (E) are schematized spectra of pitch-evoking stimuli; (F) is the subharmonic histogram of the spectrum in (E). Choosing the peak in the spectrum reveals the pitch in (A) but not in (B) where there are several peaks. Choosing the largest peak works in (B) but fails in (C). Choosing the peak with lowest frequency works in (C) but fails in (D). Choosing the spacing between peaks works in (D) but fails in (E). A pattern-matching scheme (F) works with all stimuli. The cue to pitch here is the rightmost among the largest bins (*bold line*).

is still the same. A reasonable modification is to replace the largest peak by the peak of lowest frequency, but consider now the sound illustrated in Figure 6.1D. The lowest peak is at a higher harmonic, yet the pitch is still the same. A reasonable modification is to use the spacing between partials as a measure of period. That is all the more reasonable as it often determines the frequency of the temporal envelope of the sound, as well as the frequency of possible difference tones (distortion products) resulting from nonlinear interaction between adjacent partials. However, consider now the sound illustrated in Figure 6.1E. None of the interpartial intervals corresponds to its pitch, which (for some listeners) is the same as that of the other tones.

This brings us to a final algorithm. Build a histogram in the following way: for each partial, find its subharmonics by dividing the frequency of the partial by successive small integers. For each subharmonic, increment the corresponding histogram bin. Applied to the spectrum in Figure 6.1E, this produces the histogram illustrated in Figure 6.1F. Among the bins, some are larger than the rest. The rightmost of the (infinite) set of largest bins is the cue to pitch. This

algorithm works for all the spectra shown. It illustrates the principle of *pattern matching* models of pitch perception.

## 2.2 Waveform

The waveform approach operates directly on the stimulus waveform. Consider again our pure tone, illustrated in the time domain in Figure 6.2A. Its periodic nature is obvious as a regular repetition of the waveform. A way to measure its period is to find landmarks such as peaks (shown as arrows) and measure the interval between them. This works for a pure tone, but consider now the sound in Figure 6.2B that evokes the same pitch. It has two peaks within each period, whereas our algorithm expects only one. A trivial modification is to

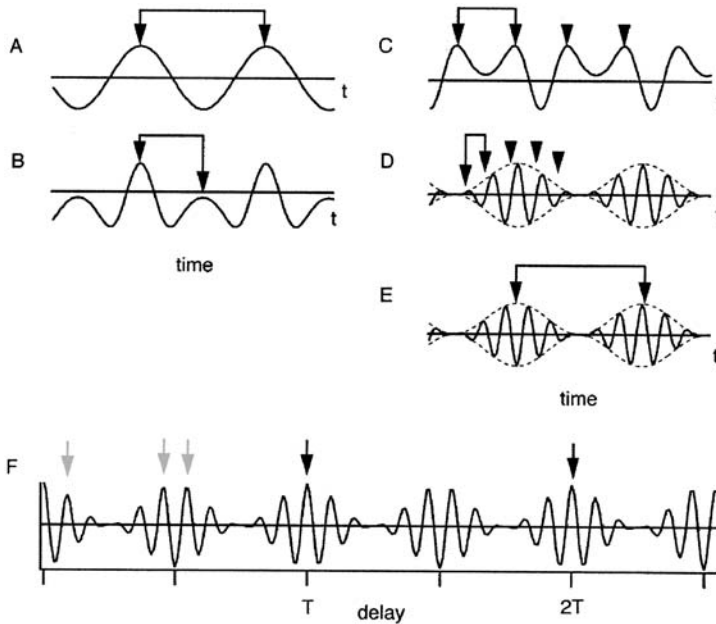


FIGURE 6.2. Temporal approach. (A) to (E) are waveform samples of pitch-evoking stimuli. (F) is the autocorrelation function of the waveform in (E). Taking the interval between successive peaks (*arrows*) works in (A) but fails in (B). The interval between highest peaks works in (B) but fails in (C). The interval between positive-going zero-crossings works in (C) but fails in (D) where there are several zero-crossings per period. The envelope works in (D), but fails in (E). A scheme based on the autocorrelation function (F) works for all stimuli. The leftmost of the (infinite) series of main peaks (*dark arrows*) indicates the period. Stimuli such as (E) tend to be ambiguous and may evoke pitches corresponding to the *gray arrows* instead of (or in addition to) the pitch corresponding to the period.

use the most prominent peak of each period, but consider now the sound in Figure 6.2C. Two peaks are equally prominent. A tentative modification is to use zero-crossings (e.g., negative-to-positive) rather than peaks, but then consider the sound in Figure 6.2D, which has the same pitch but several zero-crossings per period. Landmarks are an awkward basis for period estimation: it is hard to find a marking rule that works in every case. The waveform in Figure 6.2D has a clearly defined temporal envelope with a period that matches its pitch, but consider now the sound illustrated in Figure 6.2E. Its pitch does not match the period of its envelope (as long as the ratio of carrier to modulation frequencies is less than about 10; see Plack and Oxenham, Chapter 2).

This brings us to a final algorithm that uses, as it were, every sample as a “landmark.” Each sample is compared to every other in turn, and a count is kept of the intersample intervals for which the match is good. Comparison is done by taking the product, which tends to be large if samples  $x(t)$  and  $x(t - \tau)$  are similar, as when  $\tau$  is equal to the period  $T$ . Mathematically:

$$r(\tau) = \int x(t)x(t - \tau)dt \quad (6.1)$$

defines the *autocorrelation function*, illustrated in Figure 6.2F. For a periodic sound, the function is maximum at  $\tau = 0$ , at the period, and at all its multiples. The first of these maxima with a strictly positive abscissa can be used as a cue to the period. This algorithm is the basis of what is known as the autocorrelation (AC) model of pitch. Autocorrelation and pattern matching are both adequate to measure periods as required by a pitch model, and they form the basis of modern theories of pitch perception.

We reviewed a number of principles, some of which worked and others not. All have been used in one pitch model or another. Those that use a flawed principle can (once the flaw is recognized) be ruled out. It is harder to know what to do with the models that remain. The rest of this chapter tries to chart out their similarities and differences. The approach is in part historical, but the focus is on the future more than on the past: in what direction should we take our next step to improve our understanding of pitch?

### 2.3 What Is a Model?

An important source of disagreement between pitch models, often not explicit, is what to expect of a model. The word is used with various meanings. A very broad definition is: *a thing that represents another thing in some way that is useful*. This definition also fits other words such as *theory*, *map*, *analogue*, *metaphor*, *law*, and so forth, all of which have a place in this review. “Useful” implies that the model represents its object faithfully, and yet is somehow easier to handle and thus distinct from its object. Norbert Wiener is quoted as saying: “The best material model of a cat is another, or preferably the same, cat.” I disagree: a cat is no easier to handle than itself, and thus not a useful model. Model and world must differ. Faithfulness is not sufficient. Figure 6.3 gives an example of a model that is obviously “wrong” and yet useful.

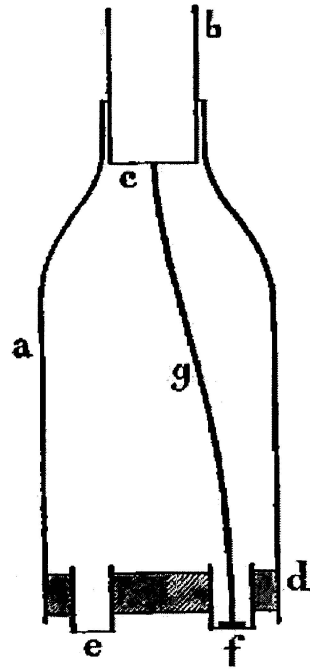


FIGURE 6.3. Johannes Müller built this model of the middle ear to convince himself that sound is transmitted from the ear drum (c) via the ossicular chain (g) to the oval window (f), rather than by air to the round window (e) as was previously thought. The model is obviously “false” (the ossicular chain is not a piece of wire) but it allowed an important advance in understanding hearing mechanisms. From Müller (1838), in von Békésy and Rosenblith (1948).

There are several corollaries. Every model is “false” in that it cannot match reality in all respects (Hebb 1959). Mismatch being allowed, multiple models may usefully serve a common reality. One pitch model may predict behavioral data quantitatively, while another is easier to explain, and a third fits physiology more closely. Criteria of quality are not one-dimensional, so models cannot always be ordered from best to worst. Rather than pit them one against another until just one (or none) remains, it is fruitful to see models as tools of which a craftsman might want several. Taking a metaphor from biology, we might argue for the “biodiversity” of models, which excludes neither competition nor the concept of “survival of the fittest.” Licklider (1959) put it this way:

The idea is simply to carry around in your head as many formulations as you can that are self-consistent and consistent with the empirical facts you know. Then, when you make an observation or read a paper, you find yourself saying, for example, “Well that certainly makes it look bad for the idea that sharpening occurs in the cochlear excitation process.”

Beginners in the field of pitch, reading of an experiment that contradicts a theory, are puzzled to find the disqualified theory live on until a new experiment contradicts its competitors. De Boer (1976) used the metaphor of the swing of a pendulum to describe such a phenomenon. An evolutionary metaphor is also fitting: as one theory reaches dominance, the others retreat to a sheltered eco-

logical niche (where they may mutate at a faster pace and emerge at a later date). This review attempts yet another metaphor, that of “genetic manipulation,” in which pieces of models (“model DNA”) are isolated so that they may be recombined, hopefully speeding the evolution of our understanding of pitch. We shall use a historical perspective to help isolate these significant strands. Before that, we need to discuss two more subjects of discord: the physical dimensions of stimuli and the psychological dimensions of pitch.

## 2.4 Stimulus Descriptions

A second source of discord is *stimulus descriptions*. There are several ways to describe and parameterize stimuli that evoke a pitch. Some fit a wide range of stimuli, others a narrower range but with some other advantage. The “best choice” depends on the problem at hand. Whatever the choice, it is important to realize that the stimulus usually differs more or less from its idealized description (one could speak of a “model” of the stimulus). We use this opportunity to introduce some notations that will be useful later on.

A first description is the *periodic* signal (Fig. 6.4A). A signal  $x(t)$  is periodic if there exists a number  $T \neq 0$  such that  $x(t) = x(t - T)$  for all time  $t$ . If there is one such number, there is an infinite set of them, and the *period* is defined as the smallest strictly positive member of that set (others are integer multiples). This representation is *parameterized* by the period  $T$  and by the shape of the waveform during a period:  $x(t)$ ,  $0 < t \leq T$ . Stimuli differ from this description in various ways: they may be of finite duration, inharmonic, modulated in frequency or amplitude, or mixed with noise, and so forth. The description is nevertheless useful: stimuli that fit it well tend to have a clear pitch that depends on  $T$ .

A second description is the *sinusoid*, defined as  $x(t) = A \cos(ft + \phi)$  where  $A$  is amplitude,  $f$  frequency and  $\phi$  the starting phase (Fig. 6.4B). A sinusoid is periodic with period  $T = 1/f$ , so this description is a special case of the previous one. Sinusoids have an additional useful property: feeding one to a *linear time-invariant* system produces a sinusoid at the output. Its amplitude is multiplied by a fixed factor and its phase is shifted by a fixed amount, but it remains a sinusoid and its frequency is still  $f$ . Many acoustic processes are linear and time invariant. This makes the sinusoid an extremely useful description.

Supposing our stimulus is almost, but not quite, sinusoidal, should we use the better-fitting periodic description, or the more tractable sinusoidal description? The advantages of the latter might make us tolerate a less good fit. Disagreement between pitch perception models can be traced, in part, to a different answer to this question.

A third way of describing a pitch-evoking stimulus is as a *sum of sinusoids*. Fourier’s theorem says that any time-limited signal may be expressed as a sum of sinusoids:

$$x(t) = \sum_k A_k \cos(2\pi f_k t - \phi_k) \quad (6.2)$$

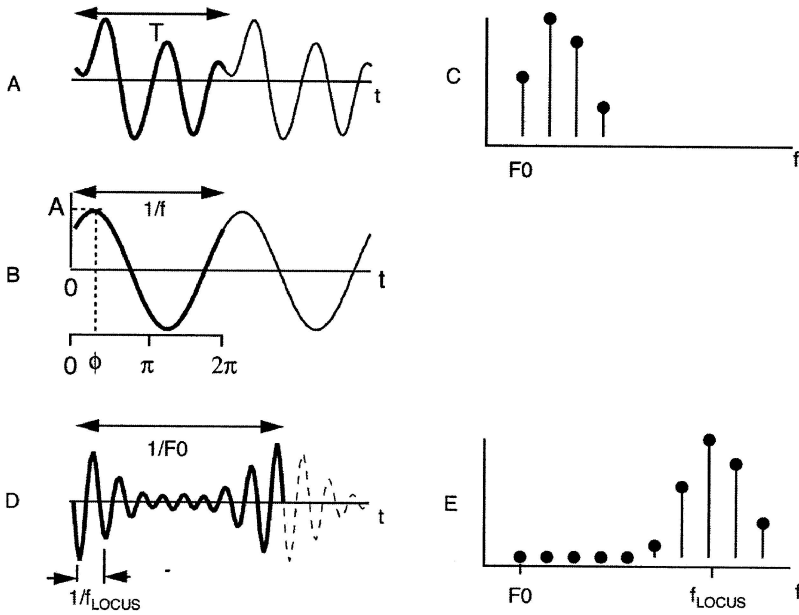


FIGURE 6.4. Descriptions of pitch-evoking stimuli. (A) Periodic waveform. The parameters of the description are  $T$  and the values of the stimulus during one period:  $s(t)$ ,  $0 < t \leq T$ . (B) Sinusoidal waveform. The parameterization ( $f$ ,  $A$ , and  $\phi$ ) is simpler, but the description fits a smaller class of stimuli (pure tones). (C) Amplitude spectrum of the signal in (A). Together with phase (not shown) this provides an alternative parameterization of the stimulus in (A). (D) Waveform of a formant-like periodic stimulus. (E) Spectrum of the same stimulus. This stimulus may evoke a pitch related to  $F_0$ , or to  $f_{\text{LOCUS}}$ , or both.

The number of terms in the sum is possibly infinite, but a nice property is that one can always select a finite subset (a “model of the model”) that fits the signal as closely as one wishes. The parameters are the set  $(f_k, A_k, \phi_k)$ . The appeal of this description is that the effect of passing the stimulus through a linear time-invariant system may be predicted from its effect on each sinusoid in the sum. It thus combines useful features of the previous two descriptions, but adds a new difficulty: each of the frequencies ( $f_k$ ) could plausibly map to pitch.

A special case is the *harmonic* complex, for which all  $(f_k)$  are integer multiples of a common frequency  $F_0$ . Parameters then reduce to  $F_0$  and  $(A_k, \phi_k)$ . Fourier’s theorem tells us that the description is now equivalent to that of a periodic signal. It fits exactly the same stimuli, and the theorem allows us to translate between parameters  $x(t)$ ,  $0 < t \leq T$  and  $(A_k, \phi_k)$ . This description fits many pitch-evoking stimuli and is very commonly used.

A fourth description is sometimes useful. The *formant* is a special case of a

sum-of-sinusoids in which amplitudes  $A_k$  are largest near some frequency  $f_{\text{LOCUS}}$  (Fig. 6.4E). Its relevance is that a stimulus that fits this model may have a pitch related to  $f_{\text{LOCUS}}$ , and if the signal is also periodic with period  $T = 1/F_0$ , pitches related to  $F_0$  and  $f_{\text{LOCUS}}$  may both be heard (some people tend to hear one more easily than the other).

These various parameterizations appear repeatedly within the history of pitch. None is “good” or “bad”: they are all tools. However, multiple stimulus parameterizations pose a problem, as parameters are the “physical” dimensions that psychophysics deals with.

## 2.5 What Is Pitch?

A third possible source of discord is the definition of pitch itself (Plack and Oxenham, Chapter 1). The American National Standard Institute defines pitch as *that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high* (ANSI 1973). It doesn’t mention the physical characteristics of the sounds. The French standards organization adds that pitch is *associated with frequency* and is low or high *according to whether this frequency is smaller or greater* (AFNOR 1977). The former definition is psychological, the latter psychophysical.

Both definitions assume a single perceptual dimension. For pure tones this makes sense, as the relevant stimulus parameter ( $f$ ) is one-dimensional. Other perceptual dimensions such as brightness might exist, but they necessarily covary with pitch (Plomp 1976). For other pitch-evoking stimuli the situation is more complex. Depending on the stimulus representation (see Section 2.4), there might be several frequency parameters. Extrapolating from the definitions, one cannot exclude the possibility of multiple pitch-like dimensions. Indeed, a stimulus that fits the “formant” signal model may evoke a pitch related to  $f_{\text{LOCUS}}$  instead of, or in addition to, the pitch related to  $F_0$ . Listeners may attend more to one or the other, and the outcome of experiments may be task and listener dependent (Smoorenburg 1970). For such stimuli, pitch has at least two dimensions, as illustrated in Figure 6.5. The pitch related to  $F_0$  is called *periodicity pitch*, and that related to  $f_{\text{LOCUS}}$  is called *spectral pitch*.<sup>1</sup> A pure tone also fits the formant model, but its periodicity and spectral pitches are not distinct (diagonal in Fig. 6.5). For other formant-like sounds they are distinct. As illustrated in Figure 6.5, periodicity pitch exists only within a limited region of the parameter space. Spectral pitch is sometimes said to be mediated by place cues, and periodicity pitch by temporal cues (see below). Spectrum and time are closely linked, however, so it is wise to reserve judgment on this point.

Periodicity pitch varies according to a linear stimulus dimension (ordinate in Fig 6.5) but it has been proposed that the perceptual structure of periodicity

---

<sup>1</sup>The term *spectral pitch* is used by Terhardt (1974) to refer to a pitch related to a resolved partial (Section 4.1, 7.2). We call that pitch a *partial pitch*.



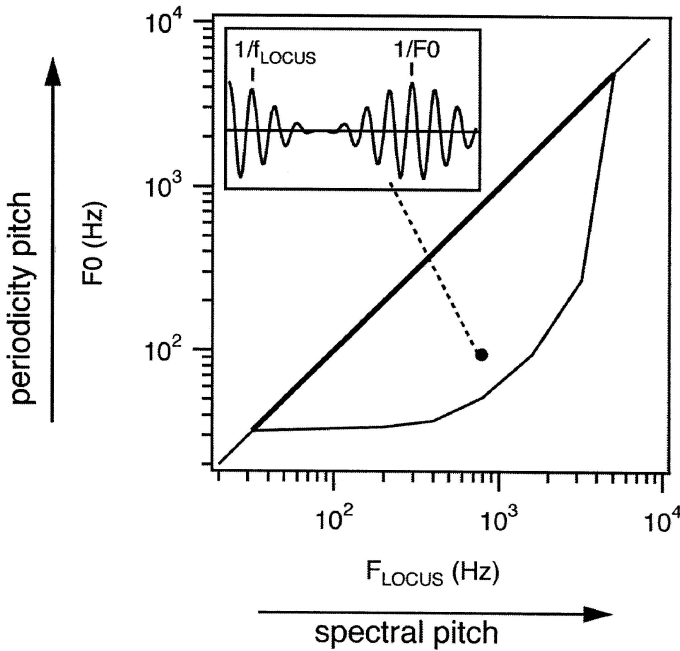


FIGURE 6.5. Formant-like stimuli may evoke two pitches, periodicity and spectral, that map to  $F_0$  and  $f_{\text{LOCUS}}$  stimulus dimensions respectively. The parameter space includes only the region below the diagonal, and stimuli that fall outside the closed region do not evoke a periodicity pitch with a musical nature (Semal and Demany 1990; Pressnitzer et al. 2001). For pure tones (diagonal) periodicity and spectral pitch covary. *Inset*: Auto-correlation function of a formant-like stimulus.

pitch is helical, with pitches distributed circularly according to *chroma* and linearly according to *tone height*. Chroma accounts for the similarity (and ease of confusion) of tones separated by an octave, and tone height for the difference between the same chroma at different octaves (Bigand and Tillmann, Chapter 9). Tone height is sometimes assumed to depend on  $f_{\text{LOCUS}}$ . However, we saw that  $f_{\text{LOCUS}}$  is a distinct stimulus dimension (abscissa in Fig. 6.5). It is the correlate of the perceptual quantity that we called spectral pitch, probably related to the dimension of brightness in timbre. Tone height and spectral pitch can be manipulated independently (Warren et al. 2003).

The pitch attribute is thus more complex than suggested by the standards, and further complexities arise as one investigates intonation in speech, or interval, melody, and harmony in music (see Bigand and Tillmann, Chapter 9). We may usefully speak of models of the pitch attribute of varying complexity. The rest of this chapter assumes the simplest model: a one-dimensional attribute related to stimulus period.

### 3. Early Roots of Place Theory

Pythagoras (6th century B.C.) is credited for relating musical intervals to ratios of string length on a monochord (Hunt 1992). The monochord is a device comprising a board with two bridges between which a string is stretched (Fig. 6.6). A third and movable bridge divides the string in two parts with equal tension but free to vibrate separately. Consonant intervals of unison, octave, fifth, and fourth arise for length ratios of 1:1, 1:2, 2:3, 3:4, respectively. This is an early example of psychophysics, in that a perceptual property (musical interval) is related to a ratio of physical quantities. It is also an early example of a model.

Aristoxenos (4th century B.C.) gives a clear, authoritative description of both interval and pitch (Macran 1902). A definition of a musical note that parallels our modern definition of pitch (ANSI 1973) was given by the Arab music theorist Safi al-Din (13th century): “a sound for which one can measure the excess of gravity or acuity with respect to another sound” (Hunt 1992). The qualitative dependency of pitch on frequency of vibration was understood by the Greeks (Lindsay 1966) but the quantitative relationship was established much later by Marin Mersenne (1636) and Galileo Galilei (1638). Mersenne proceeded in two steps. First he confirmed experimentally the laws of strings, according to which frequency varies inversely with the length of a string, proportionally to the square root of its tension, and inversely with the square root of its weight per unit length. This done, he stretched strings long enough to count the vibrations and, halving their lengths repeatedly, he derived the frequencies of every note of the scale.

Du Verney (1693) offered the first *resonance theory* of pitch perception (although the idea of resonance within the ear has earlier roots):

[The spiral lamina,] being wider at the start of the first turn than the end of the last . . . the wider parts can be caused to vibrate while the others do not . . . they are capable of slower vibrations and consequently respond to deeper tones, whereas if the narrower parts are hit, their vibrations are faster and consequently respond to sharper tones.

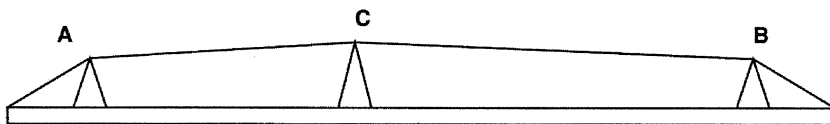


FIGURE 6.6. Monochord. A string is stretched between two fixed bridges (A, B) on a sounding board. A movable bridge (C) is placed at an intermediate position in such a way that the tension on both sides is equal. The pitches form a consonant interval if the lengths of segments AC and CB are in a simple ratio. The string plays an important role as model and metaphor in the history of pitch.

Du Verney thought that the bony spiral lamina, wide at the base and narrow at the apex, served as a resonator. Note the concept of *selective response*. He continued:

[I]n the same way as the wider parts of a steel spring vibrate slowly and respond to low tones, and the narrower parts make more frequent and faster vibrations and respond to sharp tones . . .

Du Verney used a technological metaphor to convince himself, and others, that his ideas were reasonable.

[A]ccording to the various motions of the spiral lamina, the spirits of the nerve which impregnate its substance [that of the lamina] receive different impressions that represent within the brain the various aspects of tones.

Thus was born the concept of *tonotopic projection* to the brain. This short paragraph condenses many of the concepts behind place models of pitch. The progress of anatomical knowledge up to (and beyond) Du Verney is recounted by von Békésy and Rosenblith (1948).

Mersenne was puzzled to hear, within the sound of a string or of a voice, pitches corresponding to the first five harmonics. He could not understand how a string vibrating at its fundamental could at the same time vibrate at several times that rate. He did, however, observe that a string could vibrate sympathetically to a string tuned to a multiple of its frequency, implying that it could also vibrate at that higher frequency. Simultaneity of vibration is what he could not conceive.

Sauveur (1701) observed that a string could indeed vibrate simultaneously at several harmonics (he coined the words *fundamental* and *harmonic*). The laws of strings were derived theoretically in the 18th century (in varying degrees of generality) by Taylor, Daniel Bernoulli, Lagrange, d'Alembert, and Euler (Lindsay 1966). A sophisticated theory to explain superimposed vibrations was built by Daniel Bernoulli, but Euler leap-frogged it by simply invoking the concept of *linearity*. Linearity implies the *principle of superposition*, and that is what Mersenne lacked to make sense of the several pitches he heard when he plucked a string.<sup>2</sup>

Mersenne missed the fact that the vibration he saw could reflect a sum of vibrations, with periods at integer submultiples of the fundamental period. Any such sum has the same period as the fundamental, but not necessarily the same shape. Indeed, adding sinusoidal partials produces variegated shapes depending on their amplitudes and phases ( $A_k, \phi_k$ ). That *any* periodic wave can be thus obtained, and with a *unique* set of ( $A_k, \phi_k$ ), was proved by Fourier (1822). The

---

<sup>2</sup>Mersenne pestered Descartes with this question but was not satisfied with his answers. Descartes finally came up with a qualitative explanation based on the idea of superposition in 1634 (Tannery and de Waard 1970). Superposition can be traced earlier to Leonardo da Vinci and Francis Bacon (Hunt 1992).

property had been used earlier, as many problems are solved more easily for sinusoidal movement. For example, the first derivation of the speed of sound by Newton in 1687 assumed “pendular” motion of particles (Lindsay 1966). Euler’s principle of superposition generalizes such results to any sum of sinusoids, and Fourier’s theorem adds merely that this means any waveform. This result had a tremendous impact.

## 4. Helmholtz

The mapping between pitch and period established by Mersenne and Galileo leaves a question open. An infinite number of waves have the same period: do they *all* map to the same pitch? Fourier’s theorem brings an additional twist by showing that a wave can be decomposed into elementary sinusoids. Each has its own period so, if the theorem is invoked, the period-to-pitch mapping is no longer one-to-one.

“Vibration” was commonly understood as a regular series of excursions in one direction separated by excursions in the other, but some waves have exotic shapes with several such excursion pairs per period. Do they too map to the same pitch? Seebeck (1841, in Boring 1942) found that stimuli with two or three irregularly-spaced pulses per period had a pitch that matched the period. Spacing them evenly made the pitch jump to the octave (or octave plus fifth for three pulses). In all cases the pitch was consistent with the stimulus period, regardless of shape.

Ohm (1843) objected. In his words, he had “always previously assumed that the components of a tone, whose frequency is said to be  $f$ , must retain the form  $a.\sin 2\pi ft$ .” To rescue this assumption from the results of Seebeck and others, he formulated a law saying that a tone evokes a pitch corresponding to a frequency  $f$  if and only if it “carries in itself the form  $a.\sin 2\pi(ft+p)$ .”<sup>3</sup> In other words, every sinusoidal partial evokes a pitch, and no pitch exists without a corresponding partial. In particular, periodicity pitch depends on the presence of a *fundamental partial* of nonzero amplitude. This is more restrictive than Seebeck’s condition that a stimulus merely be periodic.

Ohm’s law was attractive for two reasons. First, it drew on Fourier’s theorem, seemingly tapping its power for the benefit of hearing theory. Second, it explained the higher pitches reported by Mersenne. Paraphrasing the law, von Helmholtz (1877) stated that the sensation evoked by a pure tone is “simple” in that it does not support the perception of such higher pitches. From this he

---

<sup>3</sup>Presence of the “form” was ascertained by applying Fourier’s theorem to consecutive waveform segments of size  $1/f$ . Ohm required that  $p$  and the sign of  $a$  (but not its magnitude) be the same for each segment. He said: “The necessary impulses must follow each other in time intervals of the length  $1/f$ .” This could imply that he was referring to the pitch of the *fundamental* partial and not (as was later assumed) other partials. Authors quoting Ohm usually reformulate his law, not always with equal results.

concluded that the sensation evoked by a complex tone is composed of the sensations evoked by the pure tones it contains. A corollary is that sensation cannot depend on the relative phases of partials. This he verified experimentally for the first eight partials or so, while expressing some doubt about higher partials.

To summarize, the Ohm/Helmholtz psychoacoustic model of pitch refines the simpler law of Mersenne: (1) Among the many periodic vibrations with a given period, only those containing a nonzero fundamental partial evoke a pitch related to that period. (2) Other partials might also evoke additional pitches. (3) Relative partial amplitudes affect the quality (timbre) of the vibration, but not its pitch, as long as the amplitude of the fundamental is not zero. (4) Relative phases of partials (up to a certain rank) affect neither quality nor pitch.

The theory also included a physiological part. Sound is analyzed within the cochlea by the basilar membrane (BM) considered as a bank of radially taut strings, each loosely coupled to its neighbors. Resonant frequencies are distributed from high (base) to low (apex), and thus a sound undergoes a spectral analysis, each locus responding to partials that match its characteristic frequency. From constraints on time resolution (see Section 10.2). Helmholtz concluded that selectivity must be limited. Thus he viewed the cochlea as an approximation of the Fourier transformer needed by the psychoacoustic part of the model. Limited frequency resolution was actually welcome, as it helped him account for roughness and consonance, bringing together mathematics, physics, elementary sensation, harmony, and aesthetics into an elegant unitary theory.

Helmholtz linked the decomposition of the stimulus to a decomposition of sensation, extending the principle of superposition to the sensory domain, and to the psychoacoustic mapping between stimulus and sensation. In doing so, he assumed compositional properties of sensation and perception for which his arguments were eloquent but not quite watertight. True, his theory implies the phase-insensitivity that he observed, but to be conclusive the argument should show that it is the *only* theory that can do so. It explains Mersenne's upper pitches (each suggestive of an elementary sensation) but begs the question of why they are so rarely perceived. More seriously, it predicts something already known to be false at the time. The pitch of a periodic vibration does *not* depend on the physical presence of a fundamental partial. That was known from Seebeck's experiments, from earlier observations on beats (see Section 10.1), and from observations of contemporaries of Helmholtz cited by his translator Ellis (traduttore traditore!).

Helmholtz was aware of the problem, but argued that theory and observation could be reconciled by supposing nonlinear interaction within the ear (or within other people's sound apparatus). Distortion within the ear was accepted as an adequate explanation by later authors (von Békésy and Rosenblith 1948; Fletcher 1924) but, as Wever (1949) remarks, it does not save the psychoacoustic law. The *coup de grâce* was given by Schouten (1938), who showed that complete cancellation of the fundamental partial within the ear leaves the pitch unchanged. Licklider confirmed that that partial was dispensable by masking it, rather than

removing it. The weight of evidence against the theory as the sole explanation for pitch perception is today overwhelming (Plack and Oxenham, Chapter 2).

Nevertheless the place theory of Helmholtz is still used in at least four areas: (1) to explain pitch of pure tones (for which objections are weaker), (2) to explain the extraction of frequencies of partials (required by pattern matching theories as explained below), (3) to explain spectral pitch (associated with a spectral locus of power concentration), and (4) in textbook accounts (as a result of which the “missing fundamental” is rediscovered by each new generation). Place theory is simmering on a back burner in many of our minds.

It is tempting to try to “fix” Helmholtz’s theory retrospectively. The Fourier transform represents the stimulus according to the “sum of sinusoids” description (see Section 2.4), but among the parameters  $f_k$  of that description none is obviously related to pitch. We’d need rather an operation that fits the “periodic” or “harmonic complex” signal description. Interestingly, a string does just that. As Helmholtz (1857) himself explained, a string tuned to  $F_0$  responds to all harmonics  $kF_0$ . By superposition it responds to every sum of harmonics and therefore to any periodic sound of period  $1/F_0$  (Fig. 6.7). Helmholtz used the metaphor of a piano with dampers removed (or a harpsichord as suggested by Le Cat 1758) to explain how the ear works, and his physiological model invoked a bank of “strings” within the cochlea. However, he preferred to treat cochlear resonators as spherical resonators (which respond each essentially to a single sinusoidal component). Had he treated them as strings there would have been no need for the later introduction of pattern matching models. The “missing

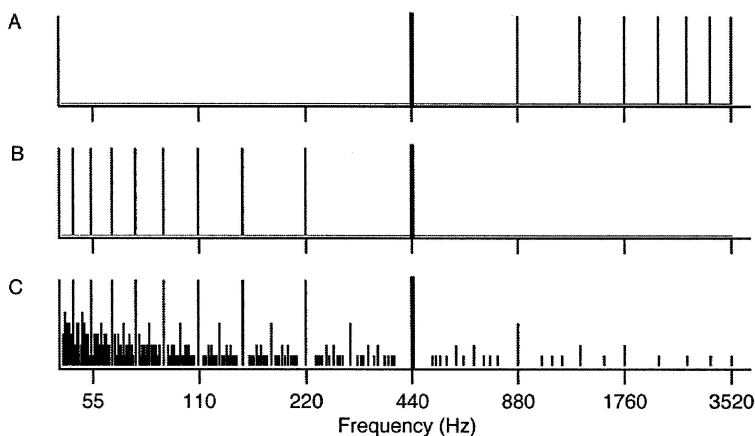


FIGURE 6.7. (A) Partial that excite a string tuned to 440 Hz. (B) Strings that respond to a 440-Hz pure tone (the abscissa of each pulse represents the frequency of the lowest mode of the string). (C) Strings that respond to a 440-Hz complex tone. Pulses are scaled in proportion to the power of the response. The rightmost string with a full response indicates the period. The string is selective to periodicity rather than Fourier frequency.

fundamental” would never be missed. Period-tuned cochlear resonators were actually suggested by Weinland in 1894 (Bonnier 1901). Of course, such a “fixed” theory holds only as long as one sees the ear as a bank of strings.

Helmholtz invoked for his theory the principle of “specific energies” of his teacher Johannes Müller, according to which each nerve represents a different *quality* (in this case a different pitch). To illustrate it, he drew on a technological metaphor: the telegraph, in which each wire transmits a single message. Alexander Graham Bell, who was trying to develop a multiplexing telegraph to overcome precisely that limitation (Hounshell 1976), read Helmholtz and, getting sidetracked, invented the telephone that later inspired to Rutherford (1886) a theory that he opposed to that of Helmholtz.

The next section shows how the missing fundamental problem was addressed by modern pitch theory.

## 5. Pattern Matching

The partials of a periodic sound form a pattern of frequencies. We are good at recognizing patterns. If they are incomplete, we tend to perceptually “reconstruct” what is missing. A pattern matching model assumes that pitch emerges in this way. Two parts are involved: one produces the *pattern* and the other looks for a match within a set of *templates*. Templates are indexed by pitch, and the one that gives the best match indicates the pitch. The best known theories are those of Goldstein (1973), Wightman (1973), and Terhardt (1974).

### 5.1 Goldstein, Wightman, and Terhardt

For Goldstein (1973) the pattern consists of a series  $f_k$  of partial frequency estimates. Each estimate is degraded by a noise, modeled as a Gaussian process with mean  $f_k$ , and a variance that is function of  $f_k$ . Only resolved partials (those that differ from their neighbors by more than a resolution limit) are included, and neither amplitudes nor phases are represented. A “central processor” attempts to account for the series as consecutive multiples of a common fundamental (the consecutiveness constraint was later lifted by Gerson and Goldstein 1978). Goldstein suggested that the  $f_k$  were possibly, but not necessarily, produced in the cochlea according to a place model such as that of Helmholtz. Srulovicz and Goldstein (1983) showed that they can also be derived from temporal patterns of auditory nerve firing. Interestingly, Goldstein mentions that estimates do not need to be ordered, and thus tonotopy need not be preserved once the estimates are known.

For Wightman (1973) the pattern consists of a tonotopic “peripheral activity pattern” produced by the cochlea, similar to a smeared power spectrum. This pattern undergoes Fourier transformation within the auditory system to produce a second pattern similar to the autocorrelation function (the Fourier transform of the power spectrum). Pitch is derived from a peak in this second pattern.

For Terhardt (1974) the pattern consists of a “specific loudness pattern” originating in the cochlea, from which is derived a pattern of partial pitches, analogous to the elementary sensations posited by Helmholtz.<sup>4</sup> From the pattern of partial pitches is derived a “gestalt” virtual pitch (periodicity pitch) via a pattern matching mechanism. Perception operates in either of two modes, analytic or synthetic, according to whether the listener accesses partial or virtual pitch, respectively. Analytic mode adheres strictly to Ohm’s law: there is a one-to-one mapping between resolved partials and partial pitches. Partial pitch is presumably innate, whereas virtual pitch is learned by exposure to speech. Listening is normally synthetic (virtual pitch).

The three models are formally similar despite differences in detail (de Boer 1977). The idea of pattern matching has roots deeper in time. It is implicit in Helmholtz’s notion of “unconscious inference” (Helmholtz 1857; Turner 1977). According to the “multicue mediation theory” of Thurlow (1963), listeners use their voice as a template (pitch then equates to the motor command that best matches an incoming sound). De Boer (1956) describes pattern matching in his thesis. Finally, pattern matching fits the behavior of the oldest metaphor in pitch theory: the *string* (compare Figs.6.1F and 6.7C).

## 5.2 Relationship to Signal Processing Methods

Signal processing methods are a source of inspiration for auditory models. Pattern matching is used in several methods of speech F0 estimation (Hess 1983). The “period histogram” of Schroeder (1968) accumulates all possible subharmonics of each partial (as in Terhardt’s model), while the “harmonic sieve” model of Duifhuis et al. (1982) tries to find a sieve that best fits the spectrum (as in Goldstein’s model). Subharmonic summation (Hermes 1988) or SPINET (Cohen et al. 1995) work similarly, and there are many variants. One is to cross correlate the spectrum with a set of “combs,” each having “teeth” at multiples of a fundamental. Rather than combs with sharp teeth, other regular patterns may be used, for example, sinusoids. Cross correlating with sinusoids implements the Fourier transform. The Fourier transform applied to a power spectrum gives the autocorrelation function (as in Wightman’s model). Applied to a logarithmic spectrum it gives the cepstrum, commonly used in speech processing (Noll 1967). There is a close connection between pattern matching and these representations.

Cochlear filters are narrow at low frequencies and wide at high. Wightman took this into account by applying nonuniform smoothing to the spectrum. Smoother parts of the spectrum require a smaller density of channels, so the spectrum can be resampled nonuniformly. This is the idea behind the so-called “mel spectrum” and MFCC (mel-frequency cepstrum coefficients), popular in speech processing. These are analogous to the logarithmic spectra of Versnel

---

<sup>4</sup>Terhardt called them *spectral pitches*, a term we reserve to designate the pitch associated with a concentration of power along the spectral axis.



and Shamma (1998). Nonuniform sampling causes the regular structure of a harmonic spectrum to be lost and thus is not very useful for pitch.

A final point is worth mentioning. We usually think of frequency as positive, but the mathematical operation that relates power spectrum to ACF (or log power spectrum to cepstrum) applies to spectra that extend over positive and negative frequencies. The negative part is obtained by reflecting the positive part over 0 Hz. Spectra are then symmetric and their Fourier spectra contain only cosines, which always have a peak at 0 Hz. A similar constraint in a harmonic comb model is to anchor a tooth at 0 Hz, and it turns out that this is important to account for the pitch of inharmonic complexes. We know that the pitch of a set of harmonics spaced by  $\Delta f$  shifts if they are all mistuned by an equal amount. Pitch varies in proportion to the central partial in a first approximation (so-called “first effect”). In a second approximation it follows a lower frequency, sometimes even lower than the lowest partial (“second effect”). Without the constraint, the best fitting comb has teeth spaced by  $\Delta f$  regardless of the mistuning, implying no pitch shift. This led Jenkins (1961) and Schouten et al. (1962) to rule out spectrum-based pattern matching models. With the constraint of a tooth anchored at 0 Hz, the best fit is a slightly stretched comb and this allows the pitch shift to be accounted for.

### 5.3 The Learning Hypothesis

Pattern matching requires a set of harmonic templates. Terhardt (1978, 1979) suggested that they are learned through exposure to harmonic-rich sounds such as speech. To explain how, Roederer (1975) proposed that spectral patterns from the cochlea are fed to a neural net. At the intersection between a channel tuned to the fundamental, and channels tuned to its harmonics, synapses are reinforced through Hebbian learning (Hebb 1949). Licklider (1959) had earlier invoked Hebbian learning to link together the period and spectrum axes of his “duplex” model. Learning was also suggested by de Boer (1956) and Thurlow (1963), and is implicit in Helmholtz’s dogma of unconscious inference (Warren and Warren 1968).

The harmonic patterns needed for learning may be found in the harmonics of a complex tone such as speech. They exist also in the series of its “superperiods” (subharmonics). This suggests that one could do away with Terhardt’s requirement of early exposure to harmonically rich sounds, since a pure tone too has superperiods. Readers in need of a metaphor to accept this idea should consider Figure 6.7. Panel A illustrates the template (made irregular by the logarithmic axis) formed by the partials of a harmonic complex tone. Panel B illustrates a similar template formed by the superperiods of a pure tone. Harmonically rich stimuli are not essential for the learning hypothesis.

Shamma and Klein (2000) went a step further and showed that template learning *does not require exposure to periodic sounds*, whether pure or complex. Their model is a significant step in the development of pattern matching models. Ingredients are: (1) an input pattern of phase locked activity, spectrally sharp or

sharpened by some neural mechanism based on synchrony, (2) a nonlinear transformation such as half-wave rectification, and (3) a matrix sensitive to spike coincidence between each channel and every other channel. In response to noise or random clicks, each channel rings at its characteristic frequency (CF). The nonlinearity creates a series of harmonics of the ringing that correlate with channels tuned to those harmonics, resulting in Hebbian reinforcement (reinforcement of a synapse by correlated activity of pre- and postsynaptic neurons) at the intersection between channels. The loci of reinforcement form diagonals across the matrix, and together these diagonals form a harmonic template. Shamma and Klein made a fourth assumption: (4) sharp phase transitions along the BM near the locus tuned to each frequency. This seems to be needed only to ensure that learning occurs also with nonrandom sounds. Shamma and Klein note that the resulting “template” is not a perfect comb. Instead it resembles somewhat Figure 6.7C.

Exposure to speech or other periodic sounds is thus unnecessary to learn a template. One can go a step further and ask whether learning itself is necessary. We noted that the string responds equally to its fundamental and to all harmonics, and thus behaves as a pattern-matcher. That behavior was certainly not learned. We’ll see later that other mechanisms (such as autocorrelation) have similar properties. Taking yet another step, we note that the string operates directly on the waveform and not on a spectral pattern. So it would seem that pattern matching itself is unnecessary, at least in terms of function. It may nevertheless be the way the auditory system works.

## 6. Pure Tones and Patterns

Pattern matching allows the response to a complex tone to be treated (in the pattern stage) as the sum of sensory responses to *pure* tones. This is fortunate, as much effort has gone into the psychophysics of pure tones. Pattern matching is not particular about how the pattern is obtained, whether by a cochlear place mechanism or centrally from temporal fine structure. It is particular about its quality: the number and accuracy of partial frequency estimates it can operate on.

### 6.1 Sharpening

Helmholtz’s estimate of cochlear resolution (about one semitone) implied that the response to a pure tone is spread over several sensory cells. Strict application of Müller’s principle would predict a “cluster” of pitches (one per cell) rather than one. Gray (1900) answered this objection by proposing that a single pitch arises at the place of *maximum stimulation*. Besides reducing the sensation to one pitch, the principle allows accuracy to be independent of peak width: narrow or wide, its locus can be determined exactly (in the absence of noise), for example by competition within a “winner-take-all” neural network (Haykin

1999). However, if noise is present before the peak is selected, accuracy obviously *does* depend on peak width. Furthermore, if two tones are present at the same time their patterns may interfere. One peak may vanish, being reduced to a “hump” on the flank of the other, or its locus may be shifted as a result of riding on the slope of the other. These problems are more severe if peaks are wide, so sharpness of the initial tonotopic pattern is important.

Recordings from the auditory nerve or the cochlea (Ruggero 1992) show tuning to be narrower than the wide patterns observed by von Békésy, which worried early theorists. Narrow cochlear tuning is explained by active mechanisms that produce negative damping. The occasional observation of spontaneous oto-acoustic emissions suggests that tuning might in some cases be arbitrarily narrow (e.g., Camalet et al. 2000), such as to sometimes cross into instability. However, these active mechanisms being nonlinear, one cannot extrapolate tuning observed with a pure tone to a combination of partials. Sharp tuning goes together with a boost of gain at the resonant frequency. The phenomenon of *suppression*, by which the response to a pure tone is suppressed by a neighboring tone, suggests that the boost (and thus the tuning) is lost if the tone is not alone. If hypersharp tuning requires that there be only one partial, it is of little use to sharpen the responses to partials a complex tone. Similar remarks apply to measures of selectivity in conditions that minimize suppression (Shera et al. 2002).

Indeed, at medium-to-high amplitudes, profiles of auditory-nerve fiber response to complex tones lack evidence of harmonic structure in cats (Sachs and Young 1979). However, profiles are better represented in the subpopulation of low-spontaneous rate fibers (see Winter, Chapter 4). Furthermore, Delgutte (1996; Cedolin and Delgutte 2005) argues that filters might be narrower in humans. Psychophysical forward masking patterns indeed show some harmonic structure (Plomp 1964). Schofner (Chapter 3) discusses the issues that arise when comparing measures between humans and animal models.

A “second filter” after the BM was a popular hypothesis before modern measurements showed sharply tuned mechanical responses. A variety of mechanisms have been put forward: mechanical sharpening (e.g., sharp tuning of the cilia or tectorial membrane, or differential tuning between tectorial and basilar membranes), sharpening in the transduction process, or sharpening by neural interaction. Huggins and Licklider (1951) list a number of schemes. They are of interest in that the question of a sharper-than-observed tuning arises repeatedly (e.g., in the template-learning model of Shamma and Klein). Some of these mechanisms might be of use also to sharpen ACF peaks (see Section 9).

Sharpening can operate on the cross-frequency profile of amplitudes, on the pattern of phases, or on both. A simple sharpening operation is an expansive nonlinearity, for example, implemented by coincidence of several neural inputs from the same point of the cochlea (on the assumption that probability of coincidence is the product of input firing probabilities). Another is spatial differentiation (more generally spatial filtering) of the amplitude pattern, for example, by summation of excitatory and inhibitory inputs of different tuning. Sharp

patterns can also be obtained using phase, for example, by transduction of the differential motion of neighboring parts within the cochlea, or by neural interaction between phase-locked responses. The lateral inhibitory network (LIN) of Shamma (1985) uses both amplitude and phase. Partial of low frequency ( $< 2$  kHz) are emphasized by phase transitions along the BM, and those of high frequency by spatial differentiation of the amplitude pattern. The hypothesis is made attractive by a recent model that uses a different form of phase-dependent interaction to account for loudness (Carney et al. 2002). In the average localized synchrony rate (ALSR) or measure (ALSM) of Young and Sachs (1979) and Delgutte (1984), a narrowband filter tuned to the characteristic frequency of each fiber measures synchrony to that frequency. The result is a pattern where partials stand out clearly. The matched filters of Srulovicz and Goldstein (1983) operate similarly. These are examples from a range of ingenious schemes to sharpen peaks of response patterns.

Alternatives to peak sharpening are to assume that a pure tone is coded by the edge of a tonotopic excitation pattern (Zwicker 1970), or that that partials of a complex tone are coded using the location of gaps between fibers responding to neighboring partials (Whitfield 1970).

## 6.2 Labeling by Synchrony

In place theory, the frequency of a partial is signaled by its position along the tonotopic axis. LIN and ALSR use phase locking merely to measure the position more finely. Troland (1930) argued that position is unreliable, and that it is better to label a channel by phase locking at the partial's frequency, an idea already put forward by Hensen in 1863 (Boring 1942). Peripheral filtering would serve merely to *resolve* partials, so that frequency can be measured and each channel labeled clearly. A nice feature of this idea is that all channels that respond to a partial contribute to characterize it (rather than just some pre-determined set). Tonotopy is not required, as noted by Goldstein (1973), but the "labels" still need to be decoded to whatever dimension underlies the harmonic templates to which the pattern is to be matched.

A possible decoder is some form of central filterbank. In the *dominant component* scheme of Delgutte (1984), each channel of the neural response is analyzed over a central filterbank, and the resulting spectral profiles combined over channels. A related principle underlies the *modulation filterbank* (e.g., Dau et al. 1996), discussed later on in the context of temporal models. An objection is that the hypothesis requires several filterbanks, one peripheral and one (or more) central. What is gained over a single filterbank? A possible answer is that transduction nonlinearity recreates the "missing fundamental" component for stimuli that lack one. However, one wonders why this is better (in terms of function) than Helmholtz's assumption of a mechanical nonlinearity preceding the cochlear filter.

From this discussion, it appears that the frequency of a pure tone (or partial) might be derived from either place *or* time cues. To decide between them,

Siebert (1968, 1970) used a simple model assuming triangle-shaped filters, nerve spike production according to a Poisson process, and optimal processing of spike trains. Calculations showed that place alone was sufficient to account for human performance. Time allowed *better* performance, and Siebert tentatively concluded that the auditory system does *not* use time. However, a reasonable form of suboptimal processing (filters matched to interspike interval histograms) gives predictions closer to behavior (Goldstein and Sruлович 1977). In a recent computational implementation of Siebert's approach, Heinz et al. (2001) found, as Siebert did, that place cues are sufficient and time cues more than sufficient to predict behavioral thresholds. However, predicted and observed thresholds were parallel for time but not for place (Fig. 6.8), and Heinz et al. tentatively concluded that the auditory system *does* use time. Interestingly, despite the severe degradation of time cues beyond 5 kHz (Johnson 1980), useful information could be exploited up to 10 kHz at least, and predicted and observed thresholds remained parallel up to the highest frequency measured, 8 kHz. Extrapolating from these results, the entire partial frequency pattern of a complex might be derived from temporal information.

To summarize, a wide range of schemes produce spectral patterns adequate for pattern matching. Some rely entirely on BM selectivity, while others ignore it. No wonder it is hard to draw the line between “place” and “time” theories! We now move on to the second major approach to pitch: time.

## 7. Early Roots of Time Theory

Boethius (Bower 1989) quotes the Greek mathematician Nicomachus (2nd century), of the Pythagorean school:

[I]t is not, he says, only one pulsation which emits a simple measure of sound; rather a string, struck only one time, makes many sounds, striking the air again and again. But since its velocity of percussion is such that one sound encompasses the other, no interval of silence is perceived, and it comes to the ears as if one pitch.

We note the idea, rooted in the Pythagorean obsession with number, that a sound is composed of several elementary sounds. Ohm and Helmholtz thought the same, but their “elements” were sinusoids. The notion of overlap between successive elementary sounds prefigures the concept of impulse response and convolution. Boethius continues:

If, therefore, the percussions of the low sounds are commensurable with the percussions of the high sounds, as in the ratios which we discussed above, then there is no doubt that this very commensuration blends together and makes one consonance of pitches.

Ratios of pulse counts play here the role later played by ratios of frequency in spectral theories. The origin of the relationship between pitch and pulse counts

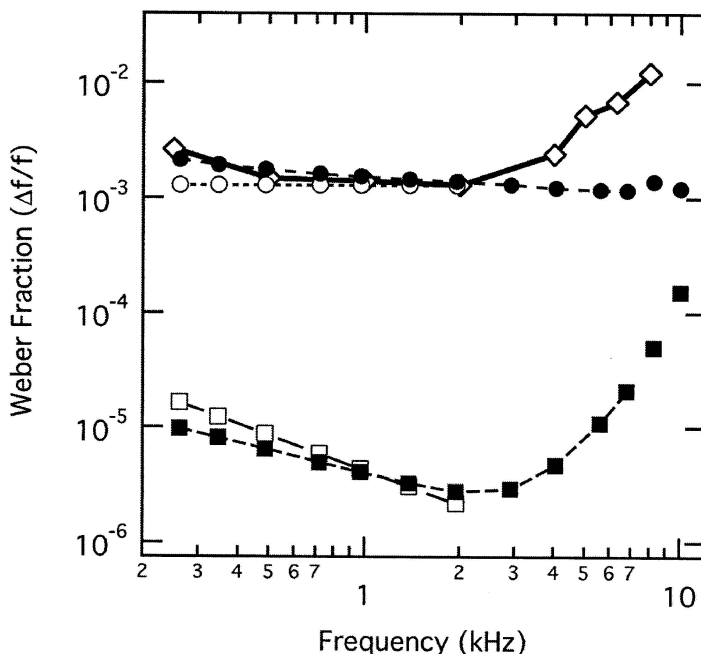


FIGURE 6.8. Pure tone frequency discrimination by humans and models, replotted from Heinz et al. (2001). *Open triangles*: Threshold for a 200-ms pure tone with equal loudness as a function of frequency (Moore 1973). *Circles*: Predictions of place-only models. *Squares*: Predictions of time-only models. *Open circles* and squares are for Siebert's (1970) analytical model, closed circles and squares are for Heinz et al.'s (2001) computational model.

is unclear, partly because the vocabulary of early thinkers (or translators, or secondary sources) did not clearly distinguish between rate of vibration, speed of propagation, amplitude of vibration, and the speed (or rate) at which one object struck another to make sound (Hunt 1992). Mersenne and Descartes clarified the roles of vibration rate and speed of propagation, finding that the former determines, while the latter is independent of, pitch. It is interesting to observe Mersenne (1636) struggle to explain this distinction using the same word ("fast") for both.

The rate–pitch relationship being established, a pitch perception model must explain how rate is measured within the listener. Mersenne and Galileo both measured vibrations by counting them, but they met with two practical difficulties: the lack of accurate time standards (Mersenne initially used his heartbeat, and in another context the time needed to say "Benedicam dominum") and the impossibility of counting fast enough the vibrations that evoke pitch. These difficulties can be circumvented by the use of calibrated resonators that we mentioned earlier on, with their own set of problems due to instability of tuning.

Here is possibly the fundamental contrast between time and place: Is it more reasonable to assume that the ear counts vibrations, or contains calibrated resonators?

This question overlaps that of *where* measurement occurs within the listener, as the ear seems devoid of counters but possibly equipped with resonators. Counting, if it occurs, occurs in the brain. The disagreement about where things happen can be traced back to Anaxagoras (5th century B.C.) for whom hearing depended simply on *penetration of sound to the brain*, and Alcmaeon of Crotona (5th century B.C.) for whom *hearing is by means of the ears, because within them is an empty space, and this empty space resounds* (Hunt 1992). The latter sentence seems to “explain” more than the first: the question is also how much “explanation” we expect of a model.

The doctrine of internal air, “*aer internus*,” had a deep influence up to the eighteenth century, when it merged gradually into the concepts of resonance and “animal spirits” (nerve activity) that eventually culminated in Helmholtz’s theory. The *telephone theory* of Rutherford (1886) was possibly a reaction against the authority of that theory (and its network of mutually supporting assumptions, some untenable such as Ohm’s law). In the minimalist spirit of Anaxagoras, Rutherford proposed that the ear merely transmits vibrations to the brain like a telephone receiver. The contrast between his modest theory (two pages), and the monumental opus of Helmholtz that it opposed, is striking. To its credit, Rutherford’s two-page theory was parsimonious, to its discredit it just shoved the problem one stage up.

An objection to the telephone theory was that nerves do not fire fast enough to follow the higher pitches. Rutherford observed transmission in a frog motor nerve up to relatively high rates (352 times per second). He did not doubt that the auditory nerve might respond faster. The need for high rates was circumvented by the *volley theory* of Wever and Bray (1930), according to which several fibers fire in turn such as to produce, together, a rate several times that of each fiber. Later measurements within fibers of the auditory nerve proved the theory wrong, in that firing is stochastic rather than regular (Galambos and Davis 1943; Tasaki 1954), but right in that fibers can indeed represent frequencies higher than their discharge rate. Steady-state discharge rates in the auditory nerve are limited to about 300 spikes per second, but the pattern of instantaneous probability can carry time structure that can be measured up to 3 to 5 kHz in the cat (Johnson 1980). The limit is lower in the guinea pig, higher in the barn owl (9 kHz, Köppl 1997), and unknown in humans.

A pure tone produces a BM motion waveform with a single peak per period, a simple pattern to which to apply the volley principle (in its probabilistic form). However, Section 2.2 showed the limits of peak-based schemes for more complex stimuli. The idea that pitch follows their temporal envelope (Fig. 6.2E), via some demodulation mechanism, was proposed by Jenkins (1961) among others. It was ruled out by the experiments of de Boer (1956) and Schouten et al. (1962) in which the partials of a modulated-carrier stimulus were mistuned by equal amounts, producing a pitch shift (as mentioned earlier). The envelope

stays the same, and this rules out not only the envelope as a cue to pitch (except for stimuli with unresolved partials; Plack and Oxenham, Chapter 2), but also interpartial spacing or difference tones. De Boer (1956) suggested that the effective cue is the spacing between peaks of the waveform fine structure closest to peaks of the envelope, and Schouten et al. (1962) pointed out that zero-crossings or other “landmarks” would work as well.

The waveform fine structure theory was criticized on several accounts, the most serious being that it predicts greater phase-sensitivity than is observed (Wightman 1973). The solution to this problem was brought by the autocorrelation (AC) model. Before moving on to that, I’ll describe an influential but confusing concept: the residue.

## 8. Schouten and the Residue

In the tradition of Boethius, Ohm and Helmholtz thought that a stimulus is composed of elements. They believed that the sensation it evokes is composed of elementary sensations, and that a one-to-one mapping exists between stimulus elements and sensory elements. The fundamental partial mapped to periodicity pitch, and higher partials to higher pitches that some people sometimes hear. Schouten (1940a) agreed to all these points but one: periodicity pitch should be mapped to a different part of the stimulus, called the *residue*. He reformulated Ohm’s law accordingly.

Schouten (1938) had confirmed Seebeck’s observation that the fundamental partial is dispensable. Manipulating individual partials of a complex with his optical siren, he trained his ear to hear them out (as Helmholtz had done before using resonators). He noted that the fundamental partial too could be heard out. The stimulus then seemed to contain *two* components with the same pitch. Introspection told him that their qualities were identical, respectively, to those of a pure tone at the fundamental and of a complex tone without a fundamental. The latter carried a salient low pitch. From his new law, Schouten reasoned that the missing-fundamental complex must either *contain* or *be* the residue. He noticed that removing additional low partials left the sharp quality intact. Low partials can be heard out, and each carries its own pitch, so Schouten reasoned that they are *not* part of the residue, whereas removing higher partials reduces the sharp quality that Schouten associated with the residue. Thus he concluded that the residue must consist of these *higher partials perceived collectively*. It somehow escaped him that periodicity pitch remains salient when the higher partials are absent.

Exclusion of resolvable partials from the residue put Schouten’s theory into trouble when it was found that they actually dominate periodicity pitch (Ritsma 1967; Plomp 1967a). Strangely enough, Schouten gave as an example a bell with characteristic tones fitting the highly resolvable series 2:3:4 (Schouten 1940b,c). Its strike note fits the missing fundamental, yet all of its partials are resolvable. De Boer (1976) amended Schouten’s definition of residue to include



all partials, which is tantamount to saying that the residue *is* the sound, rather than part of it. Schouten (1940a) had mentioned that possibility, but he rejected it as causing “a great many difficulties” without further explanation. Possibly, he believed that interaction in the cochlea between partials, strong if they are unresolved, is necessary to measure the period. The AC model (Section 9) shows that it is not.

The residue concept is no longer useful and the term “residue pitch” should be avoided. The concept survives in discussions of stimuli with “unresolved” components, commonly used in pitch experiments to ensure a complete absence of spectral cues (Section 10.4). Their pitch is relatively weak, which confirms that the residue (in Schouten’s narrow definition) is *not* a major determinant of the periodicity pitch of most stimuli.

## 9. Autocorrelation

Autocorrelation, like pattern matching, is the basis of several modern models of pitch perception. It is easiest to understand as a measure of self-similarity.

### 9.1 Self-Similarity

A simple way to detect periodicity is to take the squared difference of pairs of samples  $x(t)$ ,  $x(t - \tau)$  and smooth this measure over time to obtain a temporally stable measure of self-similarity:

$$d(\tau) = (\frac{1}{2}) \int [x(t) - x(t - \tau)]^2 dt \quad (6.3)$$

This is simply half the Euclidean distance of the signal from its time-shifted self. If the signal is periodic, the distance should be zero for a shift of one period. A relationship with the autocorrelation function or ACF (Eq. [6.1]) may be found by expanding the squared difference in Eq. 6.3. This gives the relation:

$$d(\tau) = e - r(\tau) \quad (6.4)$$

where  $e$  represents signal energy and  $r$  the autocorrelation function. Thus,  $r(\tau)$  increases where  $d(\tau)$  decreases, and peaks of one match the valleys of the other. Peaks of the ACF (or valleys of the difference function) can be used as cues to measure the period. The variable  $\tau$  is referred to as the *lag* or *delay*. The difference function  $d$  and ACF  $r$  are illustrated in Figures 6.9B and C, for the stimulus illustrated in A.

### 9.2 Licklider

Licklider (1951, 1959) proposed that autocorrelation could explain pitch. Processing occurs within the auditory nervous system, after cochlear filtering and

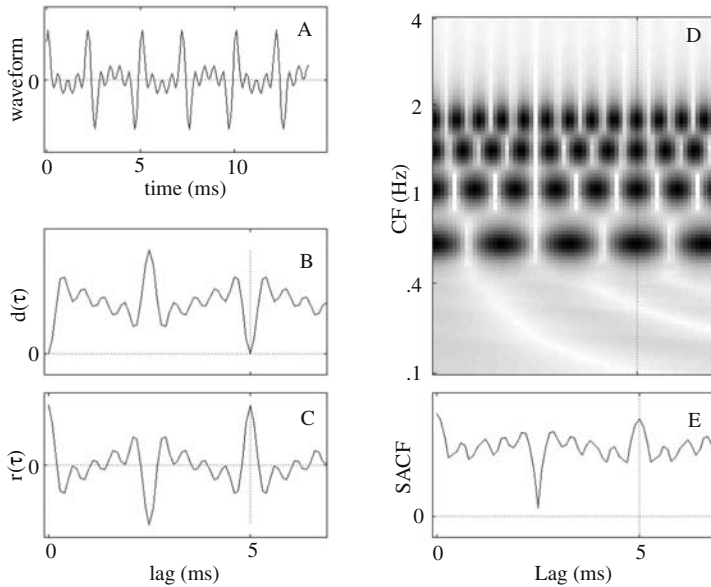


FIGURE 6.9. (A) Stimulus consisting of odd harmonics 3, 5, 7, and 9. (B) Difference function  $d(\tau)$ . (C) AC function  $r(\tau)$ . (D) Array of ACFs as in Licklider's model. (E) Summary ACF as in Meddis and Hewitt's model. Vertical dotted lines indicate the position of the period cue. Note that the partials are resolved and form well-separated horizontal bands in (D). Each band shows the period of a partial, yet their sum (E) shows the fundamental period.

hair-cell transduction. It can be modeled as operating on the half-wave rectified basilar-membrane displacement. The result is a two-dimensional pattern with dimensions characteristic frequency (CF) and lag (Fig. 6.9D). If the stimulus is periodic, a ridge spans the CF dimension at a lag equal to the period. Pitch may be derived from the position of this ridge, but Licklider didn't actually give a procedure for doing so.

Meddis and Hewitt (1991a,b) repaired this oversight by simply summing the two-dimensional pattern across frequency to produce a "summary ACF" (SACF) from which the period may be derived (Fig. 6.9E). They also included relatively realistic filter and transduction models in their implementation, and showed that the model could account for many important pitch phenomena. "AC model" in this chapter designates a class of models in the spirit of Licklider, and Meddis and Hewitt. The SACF is visually similar to the ACF of the stimulus waveform (Fig. 6.9C), which has been used as a simpler predictive model (de Boer 1956; Yost 1996).

Licklider imagined an elementary network made of neural delay elements and coincidence counters. A coincidence counter is a neuron with two excitatory synapses, that fires if spikes arrive within some short time window at both

synapses. Its firing probability is the product of firing probabilities at its inputs, and this implements the product within the formula of the ACF. Licklider supposed that this elementary network was reproduced within each channel from the periphery. It is similar to the network proposed by Jeffress (1948) to explain localization on the basis of interaural time differences.

Figure 6.9 illustrates the fact that the AC model works well with stimuli with resolved partials. Individual channels do not show fundamental periodicity (D), and yet the pattern that they form collectively is periodic at the fundamental. The period is obvious in the SACF (E). Thus, it is not necessary that partials interact on the BM to derive the period, a fact that escaped Schouten (and perhaps even Licklider himself). In the absence of half-wave rectification, the SACF would be equal to the ACF of the waveform (granted mild assumptions on the filterbank). Differences between ACF and SACF (Figs. 6.9C and E) reflect the effects of nonlinear transduction and amplitude normalization.

### 9.3 Phase Sensitivity

Excessive phase sensitivity was a major argument against temporal models (Wightman 1973). Phase refers to the parameter  $\phi$  of the sinusoid model, or  $\phi_k$  of the sum-of-sinusoids model (Section 2.4). Changing  $\phi$  is equivalent to shifting the time origin, which doesn't affect the sound. Likewise, a change of  $\phi_k$  by an amount proportional to the frequency  $f_k$  is equivalent to shifting the time origin. For a steady-state stimulus, manipulations that obey this property are imperceptible. This is de Boer's (1976) phase rule. However, phase changes that do *not* obey de Boer's rule may also be imperceptible. This is Helmholtz's rule, corollary of Ohm's law (if perception is composed from sensations, each related to a partial, there is no place for interaction *between* partials, and thus no place for phase effects). Helmholtz limited its validity to resolved partials. For stimuli with nonresolved partials, phase changes may be audible and may affect pitch, primarily the distribution of matches for ambiguous stimuli (such as illustrated in Fig. 6.2E). For example, a complex with unresolved partials in alternating sine/cosine (ALT) phase may have a pitch at the octave of its true period (Plack and Oxenham, Chapter 2).

How does the AC model fare in this respect? Autocorrelation discards phase, but it is preceded by transduction nonlinearities that *are* phase-sensitive, themselves preceded by narrow-band filters that tend on the contrary to *limit* phase-sensitive interaction. These filters are however non-linear, and they produce combination tones (see Section 10.1) that behave as extra partials with phase-dependent amplitudes.

Concretely: ACFs from channels that respond to one partial do not depend on phase (unless that partial is a phase-dependent combination tone). Channels that respond to *two* partials are only slightly phase-dependent if the partials are of high rank. Channels responding to *three* harmonics or more are more strongly phase dependent, but phase affects mainly the shape of the ACF and usually not the position of the period cue. Its *salience* may, however, change relative to

competing cues at other lags. For example, within channels responding to several partials, the ACF is sensitive to the envelope of the waveform of their sum. For complexes in ALT phase (Plack and Oxenham, Chapter 2), the envelope period is half the fundamental period, which may explain why their pitch is at the octave.

Other forms of phase sensitivity, such as to time reversal, may be accounted for by invoking a particular implementation of the AC model (de Cheveigné 1998) or related models (Patterson 1994a,b; see Section 9.5). Pressnitzer et al. (2002, 2004) describe an interesting quasi-periodic stimulus for which both the pitch and the AC model period cue *are* phase dependent. To summarize, the limited phase (in)sensitivity of the AC model accounts in large part for the limited phase (in)sensitivity of pitch (Meddis and Hewitt 1991b). See also Carlyon and Shamma (2003).

## 9.4 Histograms

Licklider's "neural autocorrelation" operation is equivalent to an all-order interspike interval (ISI) histogram, one of several formats used by physiologists to represent spike statistics of single-electrode recordings (Ruggero 1973; Evans 1986). Other common formats are first-order ISI, peristimulus time (PST), and period histograms. ISI histograms count intervals between spikes. First-order ISIs span consecutive spikes, and all-order ISIs span spikes both consecutive or not. The PST histogram counts spikes relative to the stimulus onset, and the period histogram counts them as a function of phase within the period.

Cariani and Delgutte (1996a,b) used all-order ISI histograms to quantify auditory nerve responses in the cat to a wide range of pitch-evoking stimuli. Results were consistent with the AC model. However, first-order ISI histograms are more common in the literature (e.g., Rose et al. 1967) and models similar to Licklider's have been proposed that use them (Moore 1977; van Noorden 1982). In those models, a histogram is calculated for each peripheral channel, and histograms are then summed to produce a summary histogram. The "period mode" (first large mode at nonzero lag) of the summary histogram is the cue to pitch.

Recently there has been some debate as to whether first- or all-order statistics determine pitch (Kaernbach and Demany 1998; Pressnitzer et al. 2002, 2004). Without entering the debate, we note that all-order statistics may usefully be applied to the aggregate activity of a population of  $N$  fibers. There are several reasons why one should wish to do so. One is that refractory effects prevent single fiber ISIs from being shorter than about 0.7 ms, meaning that frequencies above 800 Hz do not evoke a period mode in the first-order histogram of a single fiber. Another is that aggregate statistics make more efficient use of available information, because the number of intervals increases with the square of  $N$ . Aggregate statistics may be simulated from a single-fiber recording by pooling post-onset spike times recorded to  $N$  presentations of the same stimulus.

Intervals between spikes from the same fiber or stimulus presentation are either included (de Cheveigné 1993) or preferably excluded (Joris 2001).

In contrast, first-order statistics cannot usefully be applied to a population because, as the aggregate rate increases, most intervals join the zero-order mode (mode near zero lag, due to multiple spikes within the same period). The period mode becomes depleted, an effect accompanied by a shift of that mode towards shorter intervals (this phenomenon has actually been invoked to explain certain pitch shifts [Ohgushi 1978; Hartmann 1993]). The all-order histogram does not have this problem and is thus a better representation.

It is important to realize that any statistic discards information. Different histograms are not equivalent, and the wrong choice of histogram may lead to misleading results. For example, the ISI histogram applied to the response to certain inharmonic stimuli reveals, as expected, the “first effect of pitch shift” whereas a period histogram locked to the envelope does not (Evans 1978). Care must be exercised in the choice and interpretation of statistics.

### 9.5 Related Models

The schematic model of Moore (1977, 2003) embodies the essence of the AC model. Its description includes features (such as an upper limit on delays) that allow it to account for most important aspects of pitch (Moore 2003).

The cancellation model (de Cheveigné 1998) is based on the difference function of Eq. (6.3) instead of the ACF of Eq. (6.1). Equation 6.4 relates the two functions, and cancellation and AC models are therefore formally similar. Peaks of the ACF (Fig. 6.10A) correspond to valleys of the difference function (Fig. 6.10B). The appeal of cancellation is that it may account also for segregation of harmonic sources (de Cheveigné 1993, 1997a), which makes it useful in the context of multiple pitches (see Section 10.6). A “neural” implementation, on the lines of Licklider’s, is obtained by replacing an excitatory synapse of the coincidence neuron by an inhibitory synapse, and assuming that every excitatory spike is transmitted unless it coincides with an inhibitory spike. Roots of this model are to be found in the equalization–cancellation model of binaural interaction of Durlach (1963), and the average magnitude difference function (AMDF) method of speech F0 estimation of Ross et al. (1974) (see Hess 1983 for similar earlier methods).

The strobed temporal integration (STI) model of Patterson et al. (1992) replaces autocorrelation by cross-correlation with a train of “strobe” pulses:

$$STI(\tau) = \int s(t)x(t - \tau)dt \quad (6.5)$$

where  $s(t)$  is a train of pulses derived by some process such as peak picking. Processing occurs within each filter channel, and produces a two-dimensional pattern similar to Licklider’s. In contrast to autocorrelation, the STI operation itself is phase sensitive. It thus predicts perceptual sensitivity to time reversal of some stimuli (Patterson 1994a), although it is not clear that it also predicts

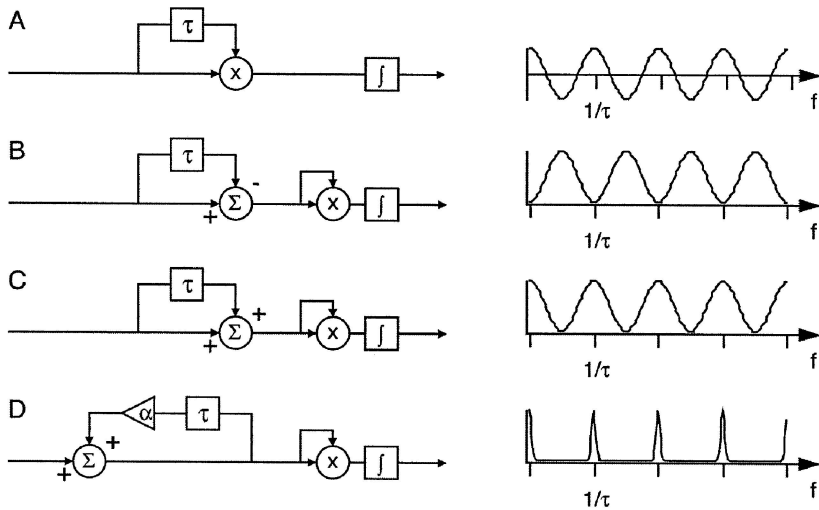


FIGURE 6.10. Processing involved in various pitch models. (A) Autocorrelation involves multiplication. (B) Cancellation involves subtraction. (C) The feed-forward comb-filter (Delgutte 1984) involves addition. (D) In the feedback comb-filter, the delayed output is added to the input (after attenuation), rather than the delayed input. This circuit behaves like a string. Plots on the right show, as a function of frequency, the value measured at the output for a pure-tone input. For a frequency inverse of the delay, and all of its harmonics, the product (A) is maximum, the difference (B) is minimum, the sum (C) is maximum. Tuning is sharper for the feedback comb-filter (D).

the insensitivity observed for others. A possible advantage of STI over the ACF is that the strobe can be delayed instead of the signal:

$$STI(\tau) = \int s(t - \tau)x(t)dt \quad (6.6)$$

in which case the implementation of the delay might be less costly (if a pulse is less expensive to delay than an arbitrary waveform). Within the brainstem, octopus cells have strobe-like properties, and their projections are well represented in man (Adams 1997). A possible weakness of STI is that it depends, as do early temporal models, on the assignment of a marker (strobe) to each period.

The term auditory image model (AIM) refers, according to context, either to STI or to a wider class including autocorrelation. Thanks to strobed integration, the fleeting patterns of transduced activity are “stabilized” to form an image. As in similar displays based on the ACF (e.g., Lyon 1984; Weintraub 1985; Slaney 1990), we can hope that visually prominent features of this image might be easily accessible to a central processor. An earlier incarnation of the image idea is the “camera acustica” model of Ewald (1898, in Wever 1949) in which the cochlea behaved as a resonant membrane. The pattern of standing waves was supposed to be characteristic of each stimulus. STI and AIM evolved from

earlier pulse ribbon and spiral detection models (Patterson and Nimmo-Smith 1986, 1987).

The dominant component representation of Delgutte (1984) and the modulation filterbank model (e.g., Dau et al. 1996) were mentioned earlier. After transduction in the cochlea, the temporal pattern within each cochlear channel is Fourier transformed, or split over a bank of internal filters, each tuned to its own “best modulation frequency” (BMF). The result is a two-dimensional pattern (cochlear CF versus modulation Fourier frequency or BMF). To the degree that this pattern resembles a power spectrum, modulation filterbank and AC models are related. The modulation filterbank was designed to explain sensitivity to slow modulations in the infrapitch range, but it has also been proposed for pitch (Wiegrefe et al. 2005).

Interestingly, the string can be seen as belonging to the AC model family. Autocorrelation involves two steps: delay and multiplication followed by temporal integration, as illustrated in Figure 6.10A. Cancellation involves delay, subtraction and squaring as illustrated in Figure 6.10B. Delgutte (1984) described a comb-filter consisting of delay, addition and (presumably) squaring as in Figure 6.10C. This last circuit can be modified as illustrated in Figure 6.10D. The frequency characteristics of both circuits have peaks at all multiples of  $f = 1/\tau$ , but the peaks of the latter are sharper. A string is, in essence, a delay line that feeds back onto itself as in Figure 6.10D. Cariani (2003) recently proposed that neural patterns might circulate within recurrent timing nets, producing a buildup of activity within loops that match the period of the pattern. This too fits the description of a string.

These examples show that autocorrelation and the string (and thus pattern matching) are closely related. They differ in the important respect of temporal resolution. At each instant, the ACF reflects a relatively short interval of its input (sum of the delay  $\tau$  and the duration of temporal smoothing). The string reflects the past waveform over a much longer interval, as information is recycled within the delay line. In effect, this allows comparisons across multiples of  $\tau$ , which improves frequency resolution at the expense of time resolution. Another way to capture regularity over longer intervals is the narrowed AC function (NAC) of Brown and Puckette (1989) in which high-order modes of the ACF are scaled and added to sharpen the period mode. The NAC was invoked by de Cheveigné (1989) and Slaney (1990) to explain acuity of pure tone discrimination. Another twist is to fit the AC histogram to exponentially-tapered “periodic templates” (Cedolin and Delgutte 2005), the best-fitting template indicating the pitch. NAC and periodic template can be seen as “subharmonic” counterparts of “harmonic” pattern-matching schemes. Once again we find strong connections between different models.

To conclude on a historical note, a precursor of autocorrelation was proposed by Hurst (1895), who suggested that sound propagates up the tympanic duct, through the helicotrema, and back down the vestibular duct. Where an ascending pulse meets a descending pulse, the BM is pressed from both sides. That position characterizes the period. More recently, Loeb et al. (1983) and Shamma

et al. (1989) invoked the BM as an alternative to neural delays. The BM is dispersive and behaves as a delay line only for a narrow-band stimulus. Delay can then be equated to phase, which brings us very close to some of the spectral sharpening schemes evoked earlier.

## 9.6 Selecting the Period Mode

The description of the AC model is not quite complete. The ACF or SACF of a periodic stimulus has several modes, one at each multiple of the period, including zero (Fig. 6.11A). The cue to pitch is the leftmost of the modes at

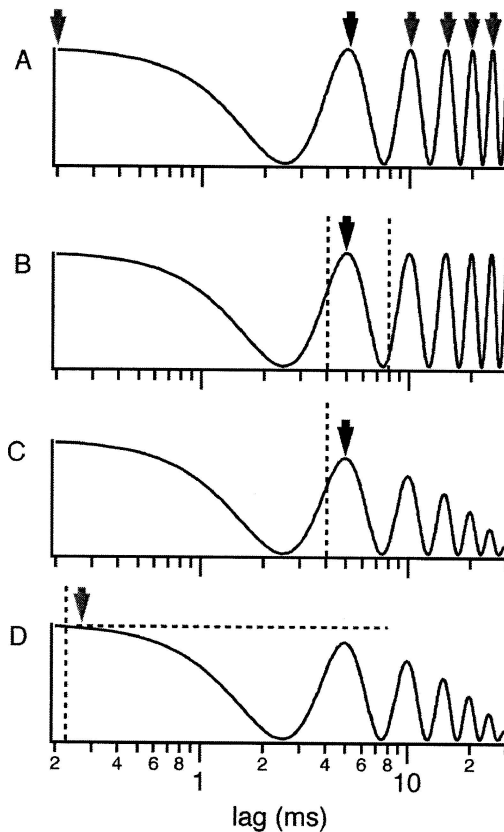


FIGURE 6.11. SACFs in response to a 200-Hz pure tone. The abscissa is logarithmic and covers roughly the range of periods that evoke a musical pitch (0.2 to 30 ms). The pitch mechanism must choose the mode that indicates the period (dark arrow in **A**) and reject the others (*gray arrows*). This may be done by setting lower and upper limits on the period range (**B**), or a lower limit and a bias to favor shorter lags. (**C**) The latter solution may fail if the period mode is less salient than the portion of the zero-lag mode that falls within the search range (**D**).



positive multiples (dark arrow). To be complete a model should specify the mechanism by which that mode is selected. A pattern-matching model is confronted with the similar problem of choosing among candidate subharmonics (Fig. 6.1F). This seemingly trivial step is one of the major difficulties in period estimation, rarely addressed in pitch models. There are several approaches.

The easiest is to set limits for the period range (Fig. 6.11B). To avoid more than one mode within the range (in which case the cue would still be ambiguous), the range must be at most one octave, a serious limitation given that musical pitch extends over about seven octaves. A second approach is to set a lower period limit and use some form of bias to favor modes at shorter lags (Fig. 6.11C). Pressnitzer et al. (2001) used such a bias (which occurs naturally when the ACF is calculated from a short-term Fourier transform, as in some implementations) to deemphasize pitch cues beyond the lower limit of musical pitch. A difficulty is that the period mode is sometimes less salient than the zero-order mode (or a spurious mode near it) (Fig. 6.11D). The difficulty can be circumvented by various heuristics, but they tend to be messy and to lack generality. A solution recently proposed in the context of F0 estimation (de Cheveigné and Kawahara 2002) is based on the difference function (Eq. [6.3], Fig. 6.9B). A normalization operation removes the dip at zero lag, after which the period lag may be selected reliably.

Once the mode (or dip) has been chosen, its position must be accurately measured. Supposing there is internal noise, it is not clear how the relatively wide modes obtained for a pure tone (Fig. 6.11) can be located with accuracy consistent with discrimination thresholds (about 0.2% at 1 kHz, Moore 1973). One solution is to suppose that higher-order modes contribute to the period estimate (e.g., de Cheveigné 1989, 2000). Another is to suppose that histograms are fed to matched filters (Goldstein and Sruлович 1977). If the task is pitch discrimination, it may not be necessary to actually choose or locate a mode. For example, Meddis and O'Mard (1997) used Euclidean distance between SACF patterns to predict discrimination thresholds. However, it is not easy to explain on that basis how a subject decides that one of two stimuli is *higher* in pitch, or how a manifold of stimuli (with same period but diverse timbres) maps to a common pitch.

To summarize, the AC model characterizes periodicity by measuring self-similarity across time, either of the acoustic waveform or of the internal patterns it gives rise to. At an abstract level, autocorrelation and pattern matching are linked via an important mathematical theorem, the Wiener–Khintchine theorem, which says that the ACF is the Fourier transform of the power spectrum. At a detailed level, they differ considerably in how they might be implemented in the auditory system. There are also important conceptual differences. For pattern matching, pure tones have the status of elementary stimuli. For the AC model they are like any other periodic stimulus, special only in that they affect a limited set of peripheral channels. Pattern matching solves the missing fundamental problem; for the AC model that problem does not occur. Pattern matching and autocorrelation, through their many variants, are the main contenders today for explaining pitch perception.

## 10. Advanced Topics

Modern pitch models account for major phenomena equally well. To decide between models, one must look at more arcane phenomena, second-order effects and implementation constraints. A model should ideally be able to fit them all; should it fail we may look to alternate models. In a sense, here is the cutting edge of pitch theory. The casual reader should skip to Section 11 and come back on a rainy day. Brave reader, read on.

### 10.1 Combination Tones

When two pure tones are added, their sum fluctuates (beats) at a rate equal to the difference of their frequencies. Young (1800) suggested that beats of the appropriate frequency could give rise to a pitch, and thus explain the “Tartini” tones sometimes observed in music (Boring 1942). By construction, the stimulus contains no partial at the beat frequency. The pitch that it evokes is therefore a counterexample to Ohm’s law.

If the medium is nonlinear, distortion products (harmonics and combination tones) may arise at the beat frequency and various other frequencies. If such were the case every time a pitch is heard, then Ohm’s law could be saved. Perhaps for that reason, there seems to have been a strong tendency to believe this hypothesis, and to assign any pitch not accounted for by a partial to a distortion product.

If the stimulus is a pure tone of frequency  $f$ , distortion products are harmonics  $nf$ . If the stimulus contains two partials at  $f$  and  $g$ , they also include terms of the form  $\pm nf \pm mg$  (where  $m$  and  $n$  are integers). Their amplitudes depend on the amplitudes of the primaries and the shape of the nonlinearity. *If the nonlinearity can be expanded as a Taylor series* around zero, these amplitudes can be calculated relatively easily (Helmholtz 1877; Hartmann 1997). The first term (linear) determines the primaries  $f$  and  $g$ . The second term (quadratic) determines the even harmonics and the difference tone  $g - f$ . The third (cubic) determines the odd harmonics and the “cubic difference tone”  $2f - g$ . Higher-order terms introduce other products. Amplitudes increase at a rate of 2 dB per dB for the difference tone, and 3 dB per dB for the cubic difference tone, as a function of the amplitude of the primaries. However all this holds only if the nonlinearity can be expanded as a Taylor series. There is no reason why that should always be the case. As a counterexample, distortion products of a half-wave rectifier vary in direct proportion to the amplitude of primaries.

The *difference tone*  $g - f$  played an important role in the early history of pitch theory. Its frequency is the same as that of beats, so it could account for the pitches that they evoke (“Tartini tones”), and also for the pitch of a “missing-fundamental” stimulus. Helmholtz argued that distortion might arise (1) within equipment used to produce “missing-fundamental” stimuli and (2) within the ear. The first argument faded with progress in instrumentation. It was already weak because periodicity pitch is salient at low amplitudes, and apparently unrelated to measurements or calculations of the difference tone.

We already noted that the second argument does not save Ohm's law, as that law claims to relate stimulus components (as opposed to internally produced) to pitches. Not only that, it is possible to cancel (and at the same time estimate) any difference tone produced by the ear, by adding an external pure tone of equal frequency, opposite phase, and appropriate amplitude (Rayleigh 1896). Adding a second low-amplitude pure tone at a slightly different frequency, and checking for the absence of beats, makes the measurement very accurate (Schouten 1938, 1970). After this very weak distortion product is canceled the pitch remains the same, so the difference tone  $g - f$  cannot account for periodicity pitch.

The harmonics  $nf$  played a confusing role. Being higher in frequency than the primaries they are expected to be more susceptible to masking than difference tones. Indeed, they are not normally perceived except at very high amplitudes. Yet Wegel and Lane (1924) found beats between a primary and a probe tone near its octave. This, they thought, indicated the presence of a relatively strong second harmonic. They estimated its amplitude by adjusting the amplitude of the probe tone to maximize the salience of beats. This method of *best beats* was widely used to estimate distortion products. Eventually, the method was found to be flawed: beats can arise from the slow variation in phase between nearly harmonically related partials (Plomp 1967b). Beats do not require closely spaced components, and thus do *not* indicate the presence of a harmonic.

This realization came after many such measurements had been published. As "proof" of nonlinearity, aural harmonics bolstered the hypothesis that the difference-tone accounts for the missing fundamental. Thus they added to confusion (on the role of difference products, see Pressnitzer and Patterson 2001). Similarly confusing were measurements of distortion products in cochlear microphonics (Newman et al. 1937), or auditory nerve-fiber responses. They arise because of nonlinear mechanical-to-nervous or electrical transduction, and do *not* reflect BM distortion components equivalent to stimulus partials, and thus are not of significance in the debate (Plomp 1965).

In contrast to other products, the *cubic difference tone*  $2f - g$  is genuinely important for pitch theory. Its amplitude varies roughly in proportion with the primaries (and not as their cube as expected from a Taylor-series nonlinearity). It increases as  $f$  and  $g$  become closer, but it is only measurable (by Rayleigh's cancellation method) for  $g/f$  ratios above 1.1, at which point it is about 14 dB below the primaries (Goldstein 1970). Amplitude decreases rapidly as the frequency spacing increases. A combination tone, even if weak, can strongly affect pitch if it falls within the dominance region (Plack and Oxenham, Chapter 2). Difference tones of higher order ( $f - n(g - f)$ ) can also contribute (Smoorenburg 1970).

Combination tones are important for pitch theory. They are necessary to explain the "second effect" of pitch shift of frequency-shifted complexes (Smoorenburg 1970; de Boer 1976). As their amplitudes are phase sensitive, they allow spectral theories to account for aspects of phase sensitivity. Their effect can be conveniently "modeled" as additional stimulus components, with

parameters that can be calculated or measured by the cancellation method (e.g., Pressnitzer and Patterson 2001). To avoid having to do so, most pitch experimenters now add low-pass noise (e.g., pink noise) to mask distortion products.

## 10.2 Temporal Integration and Resolution

A question has puzzled thinkers on and off: waves (or pulses, or particles) follow each other in time, how is it that we hear a continuous sound? Bonnier (1901), for example, argued that unipolar excitation of cochlear sensory cells would evoke an intermittent sensation if the BM did not act as a delay line (of 30 to 50 ms): at every instant, at least one cell along the delay line is excited by the excitatory phase of the waveform, allowing sensation to be continuous at least for F0s above about 20 to 30 Hz. Here we have the notion that patterns must be integrated over time to ensure smoothness (or stability of estimates over time). All models need temporal integration. It may be explicit as here, or implicit via buildup and decay of resonance.

On the other hand, Helmholtz argued that smoothing must not be excessive, because the ear needs to follow “shakes” of up to 8 notes per second that occur in music. Using  $\frac{1}{8}$  s as an upper limit on the response time of the resonators in his model, he derived a lower limit on their bandwidth, anticipating the time–frequency tradeoff of Gábor (1947) (analogous to Heisenberg’s principle of uncertainty in quantum mechanics). The tradeoff is expressed as:

$$\Delta f \Delta t \geq k \quad (6.7)$$

where  $\Delta f$  and  $\Delta t$  are frequency and time uncertainties respectively, and  $k$  is a constant that depends on how they are measured. Fine spectral resolution thus requires a long temporal analysis window. Moore (1973) calculated the resolution  $\Delta f$  with which pure tones of duration  $d$  could be discriminated on the basis of excitation pattern amplitude changes of at least 1dB. He found the relation  $\Delta f \cdot d \geq 0.24$ , analogous to Eq. (6.7). He also found that psychophysical frequency difference limens were about 10 times better than the relation implies. As Gábor’s relation is so very fundamental, this is puzzling.

The puzzle was explained by Nordmark (1968, 1970). The word “frequency” commonly carries two different meanings. One is the reciprocal of the interval between two events of equal phase, called *phase frequency* by Kneser (1948, in Nordmark 1968, 1970). The other is *group frequency* as measured by Fourier analysis:

For a time function of limited duration, [Fourier] analysis will yield a series of sine and cosine waves grouped around the phase frequency. No exact value can be given [to] the group frequency, which is thus subject to the uncertainty relation. (Nordmark, 1970)

In contrast to group frequency, phase frequency can be determined with arbitrary accuracy by measuring time between two “events.” This strong claim seems to imply the superiority of event-based (temporal) over spectral models, but we

argued earlier that events themselves are hard to extract reliably (Section 2.2). Could a similar claim be made for a model that does not use events, say, for autocorrelation?

Take an ongoing signal  $x(t)$  that is known to be periodic with some period  $T$ . Given a signal chunk of duration  $D$ , suppose that we find  $T \leq D/2$  such that  $x(t) = x(t + T)$  for every  $t$  such that both  $t$  and  $t + T$  fall within the chunk.  $T$  *might* be the period, but can we rule out other candidates  $T' \neq T$ ? Shorter periods can be ruled out by trying every  $T' \leq T$  and checking if we have  $x(t) = x(t + T')$  for every  $t$  such that both  $t$  and  $t + T'$  fall within the chunk. If this fails we can rule out a shorter period. However, we cannot rule out that the true period is longer than  $D - T$ , because our chunk might be part of a larger pattern. To rule this out we must know the longest expected period  $T_{\text{MAX}}$ , and we must have  $D \geq T + T_{\text{MAX}}$ . If this condition is satisfied, then there is no limit to the resolution with which  $T$  is determined. These conditions can be transposed to the short-term running ACF:

$$r(\tau) = \int_{t=0}^W x(t)x(t - \tau)dt \quad (6.8)$$

Two time constants are involved: the window size  $W$ , and the maximum lag  $\tau_{\text{MAX}}$  for which the function is calculated. They map to  $T_{\text{MAX}}$  and  $T$ , respectively in the previous discussion. The required duration is their sum, and depends thus on the lower limit of the expected F0 range. A rule of thumb is to allow at least  $2T_{\text{MAX}}$ .

As an example, the lower limit of melodic pitch is near 30 Hz (period  $\approx 33$  ms) (Pressnitzer et al. 2001). To estimate arbitrary pitches requires about 66 ms. If the F0 is 100 Hz (period = 1/10 ms) the time can be shortened to  $33 + 10 = 43$  ms. If we *know* that the F0 is no lower than 100 Hz, the duration may be further shortened to  $10 + 10 = 20$  ms. These estimates apply in the absence of noise. With noise present, internal or external, more time may be needed to counter its effects.

We might speculate that pattern matching allows even better temporal resolution, because periods of harmonics are shorter and require (according to the above reasoning) less time to estimate than the fundamental. Unfortunately, harmonics must be resolved, and for that the signal must be stable over the duration of the impulse response of the filterbank that resolves them.

Suppose now that the stimulus is longer than the required minimum. The extra time can be used according to at least three strategies. The first is to increase integration time to reduce noise. The second is to test for self-similarity across period multiples, so as to refine the period estimate. The third (so-called “multiple looks” strategy) is to cut the stimulus into intervals, derive an estimate from each, and average the estimates. The benefit of each can be quantified. Denoting as  $E$  the extra duration, the first strategy increases integration time by a factor  $n_1 = (E + W)/W$ , and thus reduces variability of the pattern (e.g., ACF) by a factor of  $\sqrt{n_1}$ . The second reduces variability of the *estimate* by a factor of at least  $n_2 = (E + T)/T$ , by estimating the period multiple  $n_2T$  and then dividing. It could probably do even better by including also estimates of smaller

multiples of the period. The third allows  $n_3 = (E + D)/D$  multiple looks (where  $D \geq T + W$  is interval duration), and thus reduces variability of the estimate by a factor of  $\sqrt{n_3}$ . The benefit of the first strategy is hard to judge without knowledge of the relationship between pattern variability and estimate variability. The second strategy seems better than the third (if  $n_2$  and  $n_3$  are comparable). Studies that invoke the third strategy often treat intervals as if they were surrounded by silence and thus discard structure across interval boundaries. This is certainly suboptimal. A priori, the auditory system could use any of these strategies, or some combination. The second strategy suggests a roughly inverse dependency of discrimination thresholds on duration (as observed by Moore [1973] for pure tones up to 1 to 2 kHz), while the other two imply a shallower dependency.

What parameters should be used in models? Licklider (1951) tentatively chose 2.5 ms for the size of his exponentially shaped integration windows (roughly corresponding to  $W$ ). Based on the analysis above, this size is sufficient only for periods shorter than 2.5 ms (frequencies above 250 Hz). A larger value, 10 ms, was used by Meddis and Hewitt (1992). From experimental data, Wiegrenbe et al. (1998) argued for *two* stages of integration separated by a nonlinearity. The first had a 1.5 ms window and the second some larger value. Wiegrenbe (2001) later found evidence for a period-dependent window size of about twice the stimulus period, with a minimum of 2.5 ms. These values reflect the *minimum* duration needed.

In Moore's (1973) study, pure tone thresholds varied inversely with duration up to a frequency-dependent limit (100 ms at 500 Hz), beyond which improvement was more gradual. In a task where isolated harmonics were presented one after the other in noise, Grose et al. (2002) found that they merged to evoke a fundamental pitch only if they spanned less than 210 ms. Both results suggest also a maximum integration time.

Obviously, an organism does not want to integrate for longer than is useful, especially if a longer window would include garbage. Plack and White (2000a,b) found that integration may be reset by transient events. Resetting is required by sampling models of frequency modulation (FM) or glide perception. Resetting is also required to compare intervals across time in discrimination tasks. Those tasks also require memory for the result of sampling, and it is conceivable that integration and sensory memory have a common substrate.

### 10.3 Dynamic Pitch

Aristoxenos distinguished the stationarity of a musical note, with a pitch from deep to high, from the continuity of the spoken voice or transitions between notes, with qualities of tension or relaxation. The exact terms chosen by the translator (Macran 1902) are of less interest than the fact that the concepts of static and dynamic pitch were so carefully distinguished. It is indeed conceivable that dynamic pitch is perceived differently from static pitch. For example,

FM might be transformed to amplitude modulation (AM) and perceived by an AM-sensitive mechanism (Moore and Sek 1994), or frequency glides might be decoded by a mechanism directly sensitive to the derivative of frequency (Sek and Moore 1999). The alternative is that frequency is sampled by the mechanism used for static pitch, and the samples compared across time (Hartmann and Klein 1980; Dooley and Moore 1988). For this to work, the estimation mechanism must be tolerant to frequency change.

Estimation is not instantaneous (Section 10.2), so the concept of frequency “sampling” makes sense only in a limited way. Frequency change impairs periodicity, and this makes estimation more difficult. Integration over time of unequal frequencies “blurs” the estimate of *the* frequency at any instant. A shorter window reduces the blur, but at the expense of the accuracy of the estimation process (Section 10.2).

Discrimination of frequency-modulated patterns is thus expected to be poor. Strangely, Demany and Clément (1997) observed what they called “hyperacute” discrimination of peaks of frequency modulation. Thresholds were smaller than expected given the lack of stable intervals long enough to support a sampling model. A possible explanation is that periods shrink during the upgoing ramp, and expand during the down-going ramp. Cross-period measurements that span the modulation peak are therefore relatively stable, leading to relatively good discrimination (de Cheveigné 2001).

The case might be made for the opposite proposition, that tasks involving static pitch (such as frequency discrimination) actually involve detectors sensitive to frequency change (Okada and Kashino 2003; Demany and Ramos 2004). It is often noted that weak pitches become more salient when they change (Davis 1951), so change may play a fundamental role in pitch. In the extreme one could propose that pitch is *not* a linear perceptual dimension, but rather some combination of sensitivities to pitch change and to musical interval. Whether or not this is the case, we still need to explain the extraction of the quantity that changes.

If listeners are asked to judge the overall pitch of a frequency-modulated stimulus, the result can usually be predicted from the average instantaneous frequency. If amplitude changes together with frequency, overall pitch is well predicted by the intensity- or envelope-weighted average instantaneous frequency (IWAIF or EWAIF) models (Anantharaman et al. 1993; Dai et al. 1996). Even better predictions are obtained if frequency is weighted inversely with rate of change (Gockel et al. 2001).

#### 10.4 Unresolved Partial

For Helmholtz, Ohm’s law applied only to *resolved* partials. Schouten later extended the law by assigning the remaining *unresolved* partials to a new sensory component, the residue. The resolved versus unresolved distinction is crucial for pattern matching because resolved partials alone can offer a useful pattern.

It was once crucial also for temporal models, because unresolved partials alone can produce, on the BM, the fundamental periodicity that was thought necessary for a “residue pitch.”

The distinction is still made today. Many modern studies use only stimuli with unresolved partials (to rule out “spectral cues”). Others contrast them with stimuli for which at least some partials are resolved. “Unresolved stimuli” are produced by a combination of high-pass filtering, to remove any resolved partials, and addition of low-pass noise to mask the possibly resolvable combination tones. Reasons for this interest are of two sorts. Empirically, pitch-related phenomena are surprisingly different between the two conditions (Plack and Oxenham, Chapter 2). Theoretically, pattern matching is viable only for resolved partials, so phenomena observed with unresolved partials cannot be explained by pattern matching. Autocorrelation is viable for both, but the experiments are nevertheless used to test it too. The argument is: “Autocorrelation being equally capable of handling both conditions, large differences between conditions imply that autocorrelation is *not* used for both.” The same argument applies to any unitary model. I find it not altogether convincing for two reasons: other accounts might fit the premises, and the premises themselves are not clear cut.

Auditory filters have roughly constant  $Q$ , and thus unresolved partials are necessarily of high rank. Rank, rather than resolvability, might limit performance. Indeed, Moore (2003) suggested a maximum delay of  $15/CF$  in each channel, implying a maximum rank of 15. Other possible accounts are: (1) Spectral region staying the same, unresolved stimuli must have longer periods, and longer periods may be penalized. (2) Period staying the same, unresolved stimuli must occupy higher spectral regions, and high-frequency channels might represent periodicity less well. (3) Low-pass noise added to lower spectral regions (that normally dominate pitch) in unresolved conditions may have a deleterious effect that penalizes those conditions. (4) The auditory system may learn to ignore channels where partials are unresolved, for example because they are phase sensitive (and thus more affected by reverberation), etc. These accounts need to be ruled out before effects are assigned to resolvability.

A clear behavioral difference between resolved and unresolved conditions is the order-of-magnitude step in  $F_0$  discrimination thresholds between complex tones that include lower harmonics and those that do not. The limit occurs near the 10th harmonic and is quite sharp (Houtsma and Smurzynski 1990; Shackleton and Carlyon 1994; Bernstein and Oxenham 2003). Higher thresholds are attributed to the poor resolvability of higher harmonics.

If such is the case, we expect direct measures of partial resolvability to show a breakpoint near this limit. A resolvable partial must be capable of evoking its own pitch (at least according to Terhardt’s model). An isolated partial certainly does, but two are individually perceptible only if their frequencies differ by at least 8% at 500 Hz, and somewhat more at higher or lower frequencies (Plomp 1964). Closer spacing yields a single pitch, function of the centroid of the power spectrum (Dai et al. 1996) (this justifies the assertion made in Section 2.5 that spectral pitch depends on the locus of a spectral concentration of power).



The 10th harmonic is about 9% from its closest neighbor, so this measure is roughly consistent with the breakpoint in complex F0 discrimination.

However, with neighbors on both sides, a partial is less well resolved. Harmonics in a complex are resolved only up to rank 5 to 8 (Plomp 1964). This does not agree with a breakpoint at rank 10. By pulsating the partial within the complex, Bernstein and Oxenham (2003) found a higher resolvability limit (10 to 11) that fit well with F0 discrimination thresholds in the same subjects. However, when *even* and *odd* partials were sent to different ears (thus doubling their spacing within each cochlea), partials were resolvable to about the 20th, and yet the breakpoint in F0 discrimination limens still occurred at a low rank. The two measures of resolvability do not fit.

Various other phenomena show differences between resolved and unresolved conditions: frequency modulation detection (Plack and Carlyon 1995; Carlyon et al. 2000), streaming (Grimault et al. 2000), temporal integration (Plack and Carlyon 1995; Micheyl and Carlyon 1998), pitch of concurrent harmonic sounds (Carlyon 1996), F0 discrimination between resolved and unresolved stimuli (Carlyon and Shackleton 1994; see also Oxenham et al. 2005), and so forth. If breakpoints always occurred at the same point along the resolved–unresolved continua, the resolvability hypothesis would be strengthened. However, the parameter space is often sampled too sparsely to tell. A popular stimulus set (F0s of 88 and 250 Hz and frequency regions of 125 to 625, 1375 to 1875, and 3900 to 5400 Hz) offers several resolved-unresolved continua but each is sampled only at its well-separated endpoints. Interpartial distances are drastically reduced if complex tones are added; yet “resolvability” (as defined for an isolated tone) seems to govern the salience of pitch within a mixture (Carlyon 1996). The lower limit of musical pitch increases in higher spectral regions, as expected if it was governed by resolvability, but the boundary follows a different trend, and extends well within the unresolvable zone (Pressnitzer et al. 2001). Some data do not fit the resolvable/unresolvable dichotomy.

To summarize, many modern studies focus on stimuli with unresolved partials. Aims are: (1) to test the hypothesis of distinct pitch mechanisms for resolved and unresolved complexes (Section 10.5), (2) to get more proof (if needed) that pitch can be derived from purely temporal cues, or (3) to obtain an analogue of the impoverished stimuli available to cochlear implantees (Moore and Carlyon, Chapter 7). This comes at a cost, as it focuses efforts on a region of the parameter space where pitch is weak, quite remote from the musical sounds that we usually take as pleasant. It is justified by the theoretical importance of resolvability.

### 10.5 The Two-Mechanism Hypothesis

Pattern matching and autocorrelation each has its strengths and followers. It is tempting to adopt both and assign to each a different region of parameter space: pattern matching to stimuli with resolved harmonics, and autocorrelation to stimuli with no resolved harmonics. The advantages are a better fit to data, and

better harmony between tenants of each approach. The disadvantages are that two mechanisms are involved, plus a third to integrate the two.

The temptation of multiple explanations is not new. Vibrations were once thought to take two paths through the middle ear: via ossicles to the oval window, and via air to the round window. Müller's experiment reduced them to one (Fig. 6.3). Du Verney (1683) believed that the trumpet-shaped semicircular canals were tuned like the cochlea, while Helmholtz thought the ampullae handled noise-like sounds until he realized that cochlear spectral analysis could take care of them too. Bonnier (1896–98) assigned the sacculus to sound localization (as a sort of “auditory retina”) and the cochlea to frequency analysis. Bachem (1937) postulated two independent pitch mechanisms, one devoted to tone height, the other to chroma, the latter better developed in possessors of absolute pitch. Wever (1949) suggested that low frequencies are handled by a temporal mechanism (volley theory) and high frequencies by a place mechanism, and Licklider's duplex model implemented both (with a learned neural network to connect them together). The motivation is to obtain a better fit with phenomena, and perhaps sometimes also to find a use for a component that a simpler model would ignore.

There is evidence for both temporal *and* place mechanisms (e.g., Gockel et al. 2001; Moore 2003). The assumption of independent mechanisms for resolved and unresolved harmonics is also becoming popular (Houtsma and Smurzynski 1990; Carlyon and Shackleton 1994). It has also been proposed that a unitary model might suffice (Houtsma and Smurzynski 1990; Meddis and O'Mard 1997). The issue is hard to decide. Unitary models may have serious problems (e.g., Carlyon 1998a,b) that a two-mechanism model can fix. On the other hand, assuming two mechanisms is akin to adding free parameters to a model: it automatically allows a better fit. The assumption should thus be made with reluctance (which does not mean that it is not correct). A two-mechanism model compounds vulnerabilities of both, such as lack of physiological evidence for delay lines or harmonic templates.

## 10.6 Multiple Pitches

Pitch models usually account for a single pitch, but some stimuli evoke more than one: (1) stimuli with an ambiguous periodicity pitch, (2) narrow-band stimuli that evoke both a periodicity pitch and a spectral pitch, (3) concurrent voices or instruments, and (4) complex tones in analytic listening mode.

Early experiments with stimuli containing few harmonics sometimes found multimodal distributions of pitch matches (de Boer 1956; Schouten et al. 1962). Pitch models usually produce multiple or ambiguous cues for such stimuli (e.g., Fig. 6.2F), and with appropriate weighting they should account for “multiple” pitches of this kind.

A formant-like stimulus may produce a spectral pitch related to the formant frequency (Section 2.5). The spectral pitch may coexist with a lower periodicity pitch if the stimulus is a periodic complex. For pure tones the two pitches are

confounded. In so-called diphonic singing styles of Mongolia or Tibet, spectral pitch carries the melody while periodicity pitch serves as a drone. Some listeners may be more sensitive to one or the other (Smooenburg 1970). It is common to attribute periodicity pitch to temporal analysis, and spectral pitch to cochlear analysis, reflecting two different mechanisms. However one cannot exclude a common mechanism. A sharp spectral locus implies quasi-periodicity in the time domain, and this shows up as modes at short lags in the ACF (inset in Fig. 6.5).

In music, instruments often play together, each with its own pitch, and appropriately gifted or trained people may perceive their multiple pitches (see Darwin, Chapter 8). Reverberation may transform a monodic melody into polyphony of two parts or more (the echo of a note accompanies the next). Sabine (1907) suggested that this is why scales appropriate for harmony emerged before polyphonic style. Models described so far address only the single pitch of an isolated tone, and cannot account for more without modification. A simple idea is to take the pattern that produced a pitch cue for an isolated tone, and scan it for several such cues. As an example, Assmann and Summerfield (1990) estimated the F0s of two concurrent vowels from the largest and second-largest peaks of the SACF. Unfortunately, distinct peaks do not always exist (simulations based on this procedure gave comparatively poor results; de Cheveigné 1993).

A better procedure is to estimate pitches iteratively (de Cheveigné and Kawahara 1999), by estimating first one period and then removing it. In the context of pattern matching, this is known as the “harmonic sieve” (Parsons 1976; Duifhuis et al. 1982). An initial F0 estimate is derived from the pattern of partials. Partial that fit its harmonic series (within some tolerance) are removed, and a second F0 is estimated from the remainder. The process may be iterated, each F0 controlling the sieve in turn. Scheffers (1983) tested the idea using spectral analysis similar to that of the ear, but found that F0s were rarely both estimated correctly. The reason given was lack of spectral resolution. As discussed in Section 10.4, partials within 8% to 10% of another partial are not readily resolved (they tend to merge and give rise to a single, intermediate pitch). Since many partials of a mixture have closer spacing, the applicability of a “harmonic sieve” is limited.

Iterative estimation works also with the AC model. A first period is estimated from the SACF, channels dominated by that period are discarded, and a second period is estimated from the remainder. Weintraub (1985) and Meddis and Hewitt (1992) used this procedure to segregate speech sounds. Cancellation (Section 9.5) can be used in place of autocorrelation, but it offers additional options. A period may be suppressed within a channel, for example to estimate a tone too weak to dominate any channel. The steps of suppression and estimation may also be merged into a joint estimation procedure (de Cheveigné and Kawahara 1999).

The harmonic sieve requires that partials be spaced wide enough to be resolved. Meddis and Hewitt’s scheme requires spectral envelopes, with features

(e.g., formants) broad enough to be resolved. Cancellation (if implemented perfectly) does not depend on peripheral resolution. Carlyon (1996) found that subjects could *not* perceive two pitches within pairs of “unresolved” complexes (see Section 10.4) so the effectiveness of cancellation, if used by the auditory system, must have limits.

As noted by Mersenne (1636), careful listening to a complex reveals higher pitches in addition to the fundamental. Helmholtz (1857, 1877) attributed each partial pitch to an elementary sensation produced by a sinusoidal partial.<sup>5</sup> Partial pitches are not commonly heard, but for Helmholtz they nevertheless underlie all musical perception. We access the lowest partial pitch to perceive the note, the next partial pitches to hear overtones, and the ensemble of partial pitches to hear timbre (Watt 1917 used the word “pitch-blend”). Schouten instead mapped the note to the residue, and Terhardt mapped it to the pattern of partial pitches (his “spectral pitches”), but neither disagreed with Helmholtz’s compositional model of auditory perception.

To account for partial pitches, a pattern-matching model must access the inputs of the pattern-matching stage in addition to its output (e.g., Terhardt et al. 1982; see also Martens 1984). The AC model instead accounts for them by restricting its processing to particular channels from the periphery. Helmholtz (1857) noted that partials are easier to hear out if mistuned. Mistuning also produces a systematic shift of the partial pitch (Hartmann and Doty 1996) for which an explanation, based on a time-domain process akin to the harmonic sieve, was proposed by de Cheveigné (1997b, 1999).

To summarize, there are several ways to allow pitch models to handle more than one pitch. Pattern matching models split patterns according to a “harmonic sieve” before matching. AC models divide cochlear channels among sources before periodicity estimation. Cancellation models allow joint estimation of multiple periods. For pattern matching, a partial pitch is a preexisting sensory element, perceptible if it manages to escape fusion. For AC models, it results from a segregation mechanism that involves peripheral (and possibly central) filtering. There are close relationships between pitch and segregation (Hartmann 1996; Darwin, Chapter 8). More behavioral data are needed to understand multiple pitch perception.

## 10.7 *Harmony, Melody, and Timbre*

Music science was central to science up to the 17th century. The work of Beeckmann, Descartes, Mersenne, the Galilei, and others, were largely aimed at questions such as musical consonance and musical scales (Cohen 1984). Later progress required isolating pitch from the musical context, but that context ob-

---

<sup>5</sup>Helmholtz’s translator Ellis remarked that a partial pitch might correspond instead to a series of harmonically related partials. For example, the partial pitch at the octave might correspond to the series (2, 4, 6, etc.) rather than to the 2nd harmonic, and might even exist in the absence of harmonic 2.

viously remains relevant and a pitch model should account for its effects. Chroma, intervals, harmony, tonality, or the relationship between pitch and timbre (Bigand and Tillmann, Chapter 9) are a challenge for pitch models.

*Chroma* designates a set of equivalence classes based on the octave relationship. In some cases chroma seems the dominant mode of pitch perception. For example, absolute pitch appears to involve mainly chroma (Bachem 1937; Miyazaki 1990; Ward 1999). Demany and Armand (1984) found that infants treated octave-spaced pure tones as equivalent. A spectral account of octave equivalence is that all partials of the upper tone belong to the harmonic series of the lower tone. A temporal account is that the period of the lower tone is a superperiod of the higher. In both cases the relation is not reflexive (the lower tone contains the upper tone but not vice versa) and is thus not a true equivalence. Furthermore, similar (if less close) relations exist also for ratios of 3, 5, 6, etc., for which equivalence is not usually invoked. Octave equivalence is not an obvious emergent property of pitch models.

Absolute pitch is rare. BM tuning and neural delays being relatively stable, it should be the rule rather than the exception. Relative pitch involves the potentially harder task of abstracting interval relationships between period cues along a periodotopic dimension. Some intervals involve simple numerical ratios for which coincidence between partials or subharmonics might be invoked, but accurate interval perception appears to be possible for nonsimple ratios too. Interval perception is not an obvious emergent property of pitch models.

Some aspects of harmony may be “explained” on the basis of simple ratios between period counts or partial frequencies (Rameau 1750; Helmholtz 1877; Cohen 1984). Terhardt et al. (1982, 1991) and Parncutt (1988) explain chord roots on the basis of Terhardt’s pattern-matching model. To the extent that pattern-matching models are equivalent to each other and to autocorrelation, similar accounts might be built on other pitch perception models (e.g., Meddis and Hewitt 1991a), but it is not clear how they account for the strong effects of tonal context described by Bigand and Tillmann in Chapter 9. Dependency of pitch on context or set was emphasized by de Boer (1976).

In Section 2.5 it was pointed out that certain stimuli may evoke two pitches, one dependent on periodicity, and another on the spectral locus of a concentration of power. The latter quantity also maps to a major dimension of timbre (brightness) revealed by multidimensional scaling (MDS) experiments (e.g., Marozeau et al. 2003). Historically there has been some overlap in the vocabulary and concepts used to describe pitch (e.g., “low” versus “high”) and timbre (e.g., “sharp” versus “dull”) (Boring 1942). In an MDS experiment Plomp (1970) showed that periodicity and spectral locus map to independent subjective dimensions. Tong et al. (1983) similarly found independent dimensions for place and rate of stimulation in a subject implanted with a multielectrode cochlear implant, while McKay and Carlyon (1999) found independent dimensions for carrier and modulator with a single electrode (see Moore and Carlyon, Chapter 7). As stressed by Bigand and Tillmann (Chapter 9), the musical properties of pitch must be taken into account by pitch models.

## 10.8 Binaural Effects

Binaural hearing has more than once played a key role in pitch theory. The proposal that sounds are localized on the basis of binaural time of arrival (Thompson 1882) implied that time (and not just spectrum) is represented internally. Once that is granted, a temporal account of pitch such as Rutherford's telephone theory becomes plausible. Binaural release from masking (Licklider 1948; Hirsh 1948) later had the same implication. In the "Huggins' pitch" phenomenon (Cramer and Huggins 1958), a pitch is evoked by white noise, identical at both ears apart from a narrow phase transition at a certain frequency. As there is no spectral structure at either ear, this was seen as evidence for a temporal account of pitch.

Huggins' pitch had prompted Licklider (1959) to formulate the triplex model, in which his own autocorrelation network was preceded by a network of binaural delays and coincidence counters, similar to the well-known localization model of Jeffress (1948). A favorable interaural delay was selected using Jeffress's model, and pitch was then derived using Licklider's model. The triplex model used the temporal structure at the output of the binaural coincidence network.

Jeffress's model involves multiplicative interaction of delayed patterns from both ears. Another model, the equalization–cancellation (EC) model of Durlach (1963), invoked addition or subtraction of patterns from both ears. These could also have been used to produce temporal patterns to feed the triplex model. However Durlach chose instead to use the profile of activity across CFs as a static tonotopic pattern. It turns out that many binaural phenomena, including Huggins' pitch, can be interpreted in terms of a "central spectrum," analogous to that produced monaurally by a stimulus with a structured (rather than flat) spectrum (Bilsen and Goldstein 1974; Bilsen 1977; Raatgever and Bilsen 1986). Phenomena seen earlier as evidence of a temporal mechanism were now evidence of a place mechanism situated at a central level.

In a task involving pitch perception of two-partial complexes, Houtsma and Goldstein (1972) found essentially the same performance if partials went to the same or different ears. In the latter case there is no fundamental periodicity at the periphery. They concluded that pitch cannot be mediated by a temporal mechanism and must be derived centrally from the pattern of resolved partials. These data were a major motivation for pattern matching. However, we noted earlier that Licklider's model does *not* require fundamental periodicity within a peripheral channel. It can derive the period from resolved partials, and it is but a small step to admit that they can come from both ears. Houtsma and Goldstein found that performance was no better with binaural presentation, despite the better resolution of the partials, favorable to pattern matching. Thus, their data could equally be construed as going *against* pattern matching.

An improved version of the EC model gives a good account of most binaural pitches (Culling et al. 1998a,b; Culling 2000). As in the earlier models of Durlach, or Bilsen and colleagues, it produces a tonotopic profile from which pitch cues are derived, but Akeroyd and Summerfield (2000) showed that the

temporal structure at the output of the EC stage could also be used to derive a pitch (as in the triplex model). A possible objection to that idea is that it requires two stages of time domain processing, which might be costly in terms of anatomy. However, de Cheveigné (2001) showed that the same processing may be performed as one stage. The many interactions between pitch and binaural phenomena (e.g., Carlyon et al. 2001) suggest that periodicity and binaural processing may be partly common.

### 10.9 Physiological Models

Models reviewed so far proceed by working out an account of how pitch *might* be extracted. The hope is that physiology will eventually provide support for a functionally successful model, but so far it has not obliged (Winter, Chapter 4). A strong objection to the AC model is the lack of evidence of autocorrelation patterns, or delays of the duration required (at least 30 ms). There is likewise little evidence in favor of pattern matching. A different approach is to start from known anatomy and physiology, and work towards a functional model. This seems a sound approach, as it only allows ingredients known to exist in the auditory system. Weaknesses are: (1) sparse sampling or technical difficulties may prevent the observation of an important ingredient, (2) experiment design and reporting are model driven, and in particular (3) the wrong choice of stimuli or descriptive statistics might bias model building in an unhelpful way.

The model of Langner (1981, 1998; Lagner and Schreiner 1988) tries to explain pitch and at the same time account for physiological responses to amplitude-modulated sinusoidal carriers. The basic circuit has two inputs. One is a pulse train phase-locked to the stimulus carrier (period  $\tau_c = 1/f_c$ ). The other is a strobe pulse locked to the modulation envelope (period  $\tau_m = 1/f_m$ ). The strobe triggers two parallel delay circuits that converge upon a coincidence neuron that activates if the delay difference between pathways equals the modulation period (or an integer multiple  $n_m \tau_m$  of that period). An array of such circuits covers periods in the pitch range.

The model has elements reminiscent of those of Licklider and Patterson (Section 9). A distinctive feature is the use of *two* delay circuits rather than one. One (called an “integrator” or “reductor”), accumulates carrier pulses up to some threshold and thus produces a delay (relative to the strobe) equal to an integer multiple of the carrier period ( $n_c \tau_c$ ). The other is an oscillator circuit that produces a burst of spikes triggered by the strobe, with a particular “intrinsic oscillation” period  $\tau_i$  (a small integer multiple of a synaptic delay of 0.4 ms). The circuit thus actually outputs *several* delayed spikes, all integer multiples of the oscillator period ( $n_o \tau_o$ ). Putting things together, coincidence can only occur if the “periodicity equation” is true:

$$n_m \tau_m = n_c \tau_c - n_o \tau_o$$

Since the required integers might not always exist, certain periods might be missing. From this one might predict a step-like trend of psychophysical pitch

matches, that Langner (1981) did indeed observe but that Burns (1982) failed to replicate. On the other hand, the equation allows many possible combinations of the six quantities that it involves. As a consequence, the behavior of the model is hard to analyze and compare with other models.

This example illustrates a difficulty of the physiology-driven approach. The physiological data were gathered in response to amplitude-modulated sinusoids, which don't quite fit the stimulus models of Section 2.4. Pitch varies with ( $f_c$ ,  $f_m$ ), but the parameter space is nonuniform: regions of true and approximate periodicity alternate, evoking either clear or weak and ambiguous pitch. The choice of parameters leads naturally to posit a model that extracts them in order to get at the pitch, but in this case the task is hard. In contrast, a study starting from pitch theory might have used stimuli with parameters easier to relate to pitch, and produced data conducive to a simpler model.

In a different approach, Hewitt and Meddis (1994), and more recently Wiegrebe and Meddis (2004) suggested that chopper cells in the cochlear nucleus (CN) converge on coincidence cells in the central nucleus of the inferior colliculus (ICC). Choppers tend to fire with spikes regularly spaced at their characteristic interval. Firing tends to align to stimulus transients and, if the period is close to the characteristic interval, the cell is entrained. Cells with similar properties may align to similar features and thus fire precisely at the same instant within each cycle, leading to the activation of the ICC coincidence cell. A different stimulus period would give a less orderly entrainment, and a smaller ICC output, and in this way the model is tuned. It might seem that periodicity is encoded in the highly regular interspike intervals. Actually, it is the temporal alignment of spikes across chopper cells, rather than ISI intervals within cells, that codes the pitch. A feature of this approach is the use of computational models of the auditory periphery and brainstem (Meddis 1988; Hewitt et al. 1992) to embody relevant physiological knowledge. Winter (Chapter 4) discusses physiologically based models more deeply.

## 10.10 Computer Models

Material models were once common (e.g., Fig. 6.3), but nowadays the substrate of choice is software. The many available software packages will not be reviewed, because progress is rapid and information quickly outdated, and because up-to-date tools can easily be found using search tools (or by asking practitioners in the field). The computer allows models of such a complexity that they are not easily understood (a situation that may arise also with mathematical models). The scientist is then in the uncomfortable position of requiring a second model (or metaphor) to understand the first. This is probably unavoidable, as the gap is wide between the complexity of the auditory nervous system and our limited cognitive abilities. We should nevertheless perhaps worry when a researcher treats a model as if it were as opaque as the auditory system. Special mention should be made of the sharing of software and source code. In addition to



making model production much easier, it allows models to be communicated, including those that are not easily described.

### *10.11 Other Modeling Approaches*

The ideas outlined in this subsection were chosen for their rather unusual view of neural processing of auditory patterns, and thus pitch.

Many theories invoke a spatial internal representation, for example tonotopic or periodotopic. A spatial map of pitch fits the high versus low spatial metaphor that we use for pitch, and thus gives us the feeling of “explaining” pitch. However that metaphor may be recent (Duchez 1989): the Greeks instead used words that fit their experience with stringed instruments, such as “tense” or “lax.” A different argument is that distinct pitches must map to (spatially) distinct motor neurons to allow distinct behavioral responses (Whitfield 1970). Licklider (1959) accepted the idea of a map, but questioned the need for it to be spatially ordered. The need for the map itself may also be questioned. Cariani (2001) reviews a number of alternate processing and representation schemes based on time.

Maps are usually understood as rate versus place representations, but time (of neural discharge relative to an appropriate reference) has been proposed as an alternative to rate (Thorpe et al. 1996). Maass (1998) gave formal proofs that so-called “spiking neural networks” are as powerful, and in some cases more powerful (in terms of network size for a given function), than networks based on rate. Time is a natural dimension of acoustic patterns, and its use within the auditory system makes sense. Within the auditory cortex, transient responses have been found with latencies reproducible to within a millisecond (Elhilali et al. 2005), consistent with a code in terms of spike time relative to a reference spike, itself triggered by a stimulus feature. Maass also pointed out that spiking networks allow arbitrary impulse responses to be synthesized by combining appropriately delayed excitatory and inhibitory postsynaptic potentials (EPSPs and IPSPs). Time-domain filters can thus be implemented within dendritic trees.

Barlow (1961) argued that a likely role of sensory relays is to recode incoming patterns so as to minimize the average number of spikes needed to represent them. For example, supposing the relay has  $M$  outputs, the most common input pattern would map to *no* spike, the  $M$  next-most common patterns to *one* spike on one output neuron, and so forth. Rare patterns would map to patterns with more spikes. The advantages are at least threefold. First, neural activity (and metabolic cost) is minimized, all the more so as  $M$  is large. Second, the relay extracts regularities in incoming patterns, and thus serves to characterize them. Third, reduced response to common patterns may increase sensitivity to less common events. Early relays would handle simple stimulus-related structure, and the later ones more abstract regularities. Periodicity is a candidate for early recoding, and the cancellation model (Section 9.5) actually implements it in some sense.

If Barlow's principle is valid, stimulus-related structure should give way to neural patterns that are sparse, as common patterns are coded by few spikes, and labile, as the system adjusts to the changing statistics of incoming patterns (Nelken et al. 2005). If so, stable maps of stimulus structure (tonotopy, etc.) at levels beyond brainstem and midbrain might reflect mainly irrelevant leftover structure. Barlow's principle fits well with Bayesian models of information processing (Barlow 2001).

Maass (2003) recently proposed a model of neural processing in two stages. The first performs a large number of nonlinear transformations on incoming patterns (he calls it a "liquid state machine"). The only requirement on transforms is that they be sufficiently diverse. The second stage learns linear combinations of these transforms. Theoretical analysis and simulations show that this model can efficiently learn arbitrary patterns. Transforms are, as it were, selected according to their usefulness. Networks such as Shamma and Klein's harmonic template, Licklider's autocorrelation, or cancellation, if they occurred, would be likely candidates for selection. This is an alternative form of the "learning hypothesis" (Section 5.3).

Licklider's (1951) pitch model is closely related to Jeffress's (1948) binaural model, and success of the latter (Joris et al. 1998) has bolstered the former. Recently the Jeffress model has been questioned (McAlpine et al. 2001). It assumes an array of spatially tuned channels within each cochlear frequency band, the channel with maximal activation indicating azimuth. McAlpine and colleagues instead found evidence in the guinea pig for a mechanism analogous to that which encodes color within the visual system. Azimuth affects the balance of activation of *two* channels within each frequency band, one encoding "leftness" and the other "rightness." In other words, within each cochlear frequency band, delay can be assimilated to phase and synthesized as the weighted sum of two quadrature signals. It is logical to ask if a similar mechanism could work for pitch, for example to synthesize delays required by the AC model.

Mach (1884, in Boring 1942) actually proposed a two-channel "color scheme" to code pitch height as a combination of "brightness" and "dullness," while a third channel coded "richness of timbre." Köhler (1913, in Boring 1942) used a similar idea to represent "vocality" (a quality assimilated to chroma), and Schouten (1940c) mentioned a "color" scheme to represent periodicity at each point of the basilar membrane. Helmholtz (1877) had suggested combining adjacent sensory cells to represent intermediate values of pitch, in an effort to preempt the objection that their numbers were too few to code the finer grades of pitch.

Applying a scheme analogous to McAlpine's to pitch involves difficulties of two kinds. First, except in the case of pure tones close in frequency (Dai et al. 1996), adding sounds of different pitch does not produce a sound of intermediate pitch, as when colors are mixed. Second, the requirements of pitch are harder to satisfy than localization. For a narrow band signal (such as in a cochlear channel), delay can be assimilated to phase and synthesized as the weighted sum of two signals in quadrature phase ( $\pm 90^\circ$ ). Up to 1.7 kHz (most of the range

of frequencies studied by McAlpine et al. 2001), delays of up to  $\pm 150 \mu\text{s}$  (largest guinea pig ITD) can be synthesized in this way, and if negative weights are allowed, the range can be doubled. Beyond that, the phase-delay mapping is ambiguous. The entire existence region of pitch (Fig. 6.5) involves delays longer than the period of any partial.

True, for a sufficiently narrow band signal, a large delay can be equated to phase and implemented as a delay shorter than the period (or as the weighted sum of quadrature signals). However this mapping is ambiguous and is hard to see how a pitch model can be built in this way. Nevertheless there may be some way to formulate a model along these lines that works. Certainly the need for a high-resolution array of pitch-sensitive channels might be alleviated, as originally suggested by Helmholtz.

Du Verney (1683) proposed that the eardrum is actively tuned by muscles of the middle ear to match the pitch of incoming tones (he did not say how the tunable eardrum and fixed cochlear resonators might share roles). Most pitch models are of the “fixed” sort, but tuning is possibly an option. Perception often involves some form of action, for example moving one’s head to resolve localization ambiguity. Efferent pathways are as ubiquitous within the auditory system as their role is little known (Sahey et al. 1997), and it is conceivable that pitch is extracted according to a tunable version of, say, the AC model. It might be cheaper, in terms of neural circuitry, to have one or more tunable delay/coincidence elements rather than the full array posited by the standard AC model. Tuning might explain the common lack of absolute pitch (absolute pitch would then be explained by the uncommon presence of fixed tuned elements).

To summarize Section 10, specialized issues give insight as to which model of pitch is correct, as simpler phenomena are explained equally well by most models. Special phenomena may sometimes require specialized models, but it should be understood that they all address facets of the same object, the auditory system. Hopefully some day they will merge into a unitary model worthy of Helmholtz.

## 11. Of Models and Men

This book is about pitch, but the hero of the chapter is the model. Model-making itself is a metaphor of perception. Like the shadows on the back of Plato’s cave, models reflect the world outside (or in our case: inside the ear) in the same way as the pattern of activity on the retina reflects the structure of a scene. Perception guides *action*, and effective action leads to survival of the organism. Reversing the metaphor, a criterion for judging our models is what we *do* with them. For society, the bottom line is to adequately address technical, economical, medical, and other issues. For the researcher it is to “publish or perish.” Ultimately, here is the meaning of the word “useful” in our definition of the model.

Over the past, pitch theory has progressed unevenly. Various factors appear

to have hastened or slowed the pace. Models are made by people, who are driven by whims and animosities and the need to “survive” scientifically. Ego-involvement (to use Licklider’s words) drives the model-maker to move forward, and also to thwart competition. At times, progress is fueled by the intellectual power of one person, such as Helmholtz. At others, it seems hampered by the authority of that same power. Controversy is stimulating, but it tends to lock opponents into sterile positions that slow their progress (Boring 1929, 1942).

Certain desirable features make a model fragile. A model that is specific about its implementation is more likely to be proven false than one that is vague. A model that is unitary or simple is more likely to fail than one that is narrow in scope or rich in parameters. These forces should be compensated, and at times it may be necessary to protect a model from criticism. It is my speculation that Helmholtz knew the weakness of his theory in respect to the missing fundamental, but felt it necessary to resist criticism that might have led to its demise. The value and beauty of his monumental bridge across mathematics, physiology and music were such that its flaws were better ignored. To that one must agree. Yet Helmholtz’s theory has cast a long shadow across time, still felt today and not entirely beneficial.

This chapter was built on the assumption that a healthy menagerie of models is desirable. Otherwise, writing sympathetically about them would have been much harder. There are those who believe that theories are not entirely a good thing. Von Békésy and Rosenblith (1948) expressed scorn for them, and stressed instead anatomical investigation (and technical progress in instrumentation for that purpose) as a motor of progress. Wever (1949), translator of the model-maker von Békésy, distrusted material and mathematical models. Boring (1926) called out for “fewer theories and more theorizing.” Good theories are falsifiable, and some put their best efforts into falsifying them. If, as Hebb (1959) suggests, every theory is already false by essence, such efforts are guaranteed to succeed. The falsifiability criterion is perhaps less useful than it seems.

On the other hand, progress in science has been largely a process of weeding out theories. The appropriate attitude may be a question of balance, or of a judicious alternation between the two attitudes, as in de Boer’s metaphor of the pendulum. This chapter swings in a model-sympathetic direction, future chapters may more usefully swing the other way.

Inadequate terminology is an obstacle to progress. The lack of a word, or worse, the sharing of a word between concepts that should be distinct is a source of fruitless argument. Mersenne was hindered by the need to apply the same word (“fast”) to both vibration rate and propagation speed. Today, “frequency” is associated with spectrum (and thus place theory) in some contexts, and rate (and thus temporal theory) in others. “Spectral pitch” and “residue” are used differently by different authors. We must recognize these obstacles.

Metaphors are useful. Our experience of resonating objects (Du Verney’s steel spring, or Le Cat’s harpsichord) makes the idea of resonance within the ear easy to grasp and convey to others. In this review the metaphor of the string has served to bridge time (from Pythagoras to Helmholtz to today) and theory (from

place to autocorrelation). Helmholtz used the telegraph to convince himself of the adequacy of his version of Müller's principle, but, had it been invented earlier, the telephone might have convinced him otherwise.

A final point has to do with the collective dimension of theory making. Mersenne was known to be impatient with his opponents. In 1634, Nicolas-Claude Fabri de Pieresc warned him, "You must refrain from putting criticism on others . . . without urgent necessity, to induce no one to try to bite you in revenge." Mersenne changed radically, became affable and developed an intense correspondence with the best minds of the time. In an age without scientific journals, that did possibly more for the advancement of knowledge than his own discoveries and inventions (Tannery and de Waard 1970).

## 12. Summary

Historically, theories of pitch were often theories of hearing. It is good to keep in mind this wider scope. Pitch determines the survival of a professional musician today, but the ears of our ancestors were shaped for a wider range of tasks. It is conceivable that pitch grew out of a mechanism that evolved for other purposes, for example to segregate sources, or to factor redundancy within an acoustic scene (Hartmann 1996). The "wetware" used for pitch certainly serves other functions, and thus advances in understanding pitch benefit our knowledge of hearing in general.

Ideally, understanding pitch should involve choosing, from a number of plausible mechanisms, the one used by the auditory system, on the basis of available anatomical, physiological or behavioral data. Actually, many schemes reviewed in Sections 2.1 and 2.2 were functionally weak. Understanding pitch also involves weeding out those schemes that "do not work," which is all the more difficult as they may seem to work perfectly for certain classes of stimuli. Two schemes (or families of schemes) are functionally adequate: pattern matching and autocorrelation. They are closely related, which is hardly surprising as they both perform the same function: period estimation. For that reason it is hard to choose between them.

My preference goes to the autocorrelation family, and more precisely to cancellation (that uses minima rather than maxima as cues to pitch, Section 9.5). This has little to do with pitch, and more with the fact that cancellation is useful for segregation and fits the ideas on redundancy-reduction of Barlow (1961). I am also, as Licklider put it, "ego involved." Cancellation could be used to measure periods of resolved partials in a pattern-matching model, but the pattern-matching part would still need accounting for. A period-sized delay seems an easy way to implement a harmonic template or sieve. Although the existence of adequate delays is controversial, they are a reasonable requirement compared to other schemes. If a better scheme were found to enforce harmonic relations, I'd readily switch from autocorrelation/cancellation to pattern matching. For now, I try to keep both in my mind as recommended by Licklider.

It is conceivable that the auditory system uses neither. A reason to believe so is that they don't seem to fit with every feature described by the physiologist, the psychoacoustician or the musician. Another is that both models were designed to be simple and easily understood. Obviously the auditory nervous system has no such constraint, so the actual mechanism might be far more complex than we can easily apprehend. Our current models may still be useful as tools to understand such a complex mechanism. Judging from yesterday's progress, however, it is wise to assume that yet better tools are to come.

This chapter reviewed models, present and past. Not to write a history, nor to select the best of today's models, but rather to help with the development of future models. To quote Flourens (Boring 1963): "Science *is* not. It becomes."

### 13. Sources

Delightful introductions to pitch theory (unfortunately hard to find) are Schouten (1970) and de Boer (1976). Plomp gives historical reviews on resolvability (Plomp 1964), beats and combination tones (Plomp 1965, 1967b), consonance (Plomp and Levelt 1965), and pitch theory (Plomp 1967a). The early history of acoustics is recounted by Hunt (1992), Lindsay (1966), and Schubert (1978). Important early sources are reproduced in Lindsay (1973) and Schubert (1979). The review of von Békésy and Rosenblith (1948) is oriented towards physiology. Wever (1949) reviews the many early theories of cochlear function, earlier reviewed by Watt (1917), and yet earlier by Bonnier (1896–98, 1901). Boring (1942) provides an erudite and in-depth review of the history of ideas in hearing and the other senses. Cohen (1984) reviews the progress in musical science in the critical period around 1600. Turner (1977) is a source on the Seebeck/Ohm/Helmholtz dispute. Original sources were consulted whenever possible, otherwise the secondary source is cited. For lack of linguistic competence, sources in German (and Latin for early sources) are missing. This constitutes an important gap.

*Acknowledgements.* I thank the many people who offered ideas, comments or criticism on earlier drafts, in particular Yves Cazals, Laurent Demany, Richard Fay, Bill Hartmann, Stephen McAdams, Ray Meddis, Brian Moore, Andrew Oxenham, Chris Plack, Daniel Pressnitzer, and François Raveau. Michael Heinz kindly provided data for Figure 6.8.

### References

- Adams JC (1997) Projections from octopus cells of the posteroventral cochlear nucleus to the ventral nucleus of the lateral lemniscus in cat and human. *Audit Neurosci* 3: 335–350.
- AFNOR (1977) Recueil des normes françaises de l'acoustique. Tome 1 (vocalulaire), NFS30–107. Paris: Association Française de Normalisation.

- Akeroyd MA, Summerfield AQ (2000) A fully-temporal account of the perception of dichotic pitches. *Br J Audiol* 33:106–107.
- Anantharaman JN, Krishnamurti AK, and Feth LL (1993) Intensity weighting of average instantaneous frequency as a model of frequency discrimination. *J Acoust Soc Am* 94:723–729.
- ANSI (1973) American national psychoacoustical terminology-S3.20. New York: American National Standards Institute.
- Assmann PF, Summerfield Q (1990) Modeling the perception of concurrent vowels: vowels with different fundamental frequencies. *J Acoust Soc Am* 88:680–697.
- Bachem A (1937) Various kinds of absolute pitch. *J Acoust Soc Am* 9:145–151.
- Barlow HB (1961) Possible principles underlying the transformations of sensory messages. In: Rosenblith WA (ed), *Sensory Communication*. Cambridge, MA: MIT Press, pp. 217–234.
- Barlow HB (2001) Redundancy reduction revisited. *Network Comput. Neural Syst* 12: 241–253.
- Bernstein JG, Oxenham A (2003) Pitch discrimination of diotic and dichotic tone complexes: harmonic resolvability or harmonic number? *J Acoust Soc Am* 113:3323–3334.
- Bilsen FA (1977) Pitch of noise signals: evidence for a “central spectrum”. *J Acoust Soc Am* 61:150–161.
- Bilsen FA, Goldstein JL (1974) Pitch of dichotically delayed noise and its possible spectral basis. *J Acoust Soc Am* 55:292–296.
- Bonnier P (1896–98) *L’oreille — Physiologie — Les fonctions*. Paris: Masson et fils Gauthier-Villars et fils.
- Bonnier P (1901) *L’audition*. Paris: Octave Doin.
- Boring EG (1926) Auditory theory with special reference to intensity, volume and localization. *Am J Psychol* 37:157–188.
- Boring EG (1929) The psychology of controversy. *Psychol Rev* 36:97–121 (reproduced in Boring 1963).
- Boring EG (1942) *Sensation and Perception in the History of Experimental Psychology*. New York: Appleton-Century-Crofts.
- Boring EG (1963) *History, Psychology and Science* (Edited by R.I. Watson and D.T. Campbell). New York: John Wiley & Sons.
- Bower CM (1989) *Fundamentals of Music* (translation of *De Institutione Musica*, Anicius Manlius Severinus Boethius, d524). New Haven: Yale University Press.
- Brown JC, Puckette MS (1989) Calculation of a “narrowed” autocorrelation function. *J Acoust Soc Am* 85:1595–1601.
- Burns E (1982) A quantal effect of pitch shift? *J Acoust Soc Am* 72:S43.
- Camalet S, Duke T, Jülicher F, Prost J (2000) Auditory sensitivity provided by self-tuned critical oscillations of hair cells. *Proc Natl Acad Sci USA* 97:3183–3188.
- Cariani PA (2001) Neural timing nets. *Neural Networks* 14:737–753.
- Cariani PA (2003) Recurrent timing nets for auditory scene analysis. *Proc IEEE IJCNN*, pp. 1575–1580.
- Cariani PA, Delgutte B (1996a) Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *J Neurophysiol* 76:1698–1716.
- Cariani PA, Delgutte B (1996b) Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase-invariance, pitch circularity, rate-pitch and the dominance region for pitch. *J Neurophysiol* 76:1717–1734.
- Carlyon RP (1996) Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker. *J Acoust Soc Am* 99:517–524.

- Carlyon RP (1998a) The effects of resolvability on the encoding of fundamental frequency by the auditory system. In: Palmer A, Rees A, Summerfield AQ, Meddis R (eds), *Psychophysical and Physiological Advances in Hearing*. London: Whurr, pp. 246–254.
- Carlyon RP (1998b) Comments on “A unitary model of pitch perception” [*J Acoust Soc Am* 102, 1811–1820 (1997)]. *J Acoust Soc Am* 104:1118–1121.
- Carlyon RP, Shackleton TM (1994) Comparing the fundamental frequencies of resolved and unresolved harmonics: evidence for two pitch mechanisms? *J Acoust Soc Am* 95: 3541–3554.
- Carlyon RP, Shamma S (2003) An account of monaural phase sensitivity. *J Acoust Soc Am* 114:333–348.
- Carlyon RP, Moore BCJ, Micheyl C (2000) The effect of modulation rate on the detection of frequency modulation and mistuning of complex tones. *J Acoust Soc Am* 108: 304–315.
- Carlyon RP, Demany L, Deeks J (2001) Temporal pitch perception and the binaural system. *J Acoust Soc Am* 109:686–700.
- Carney LH, Heinz MG, Evilsizer ME, Gilkey RH, Colburn HS (2002) Auditory phase opponency: a temporal model for masked detection at low frequencies. *Acta Acustica* 88:334–347.
- Cedolin L, Delgutte B (2005) Representations of the pitch of complex tones in the auditory nerve. In: Pressnitzer D, de Cheveigné A, McAdams S, Collet L (eds), *Auditory Signal Processing: Psychophysics, Physiology and Modeling*. New York: Springer, pp. 107–116.
- Cohen HF (1984) *Quantifying Music*. Dordrecht: D. Reidel (Kluwer).
- Cohen MA, Grossberg S, Wyse LL (1995) A spectral network model of pitch perception. *J Acoust Soc Am* 98:862–879.
- Cramer EM, Huggins WH (1958) Creation of pitch through binaural interaction. *J Acoust Soc Am* 30:413–417.
- Culling JF (2000) Dichotic pitches as illusions of binaural unmasking. III. The existence region of the Fourcin pitch. *J Acoust Soc Am* 107:2201–2208.
- Culling JF, Marshall D, Summerfield Q (1998a) Dichotic pitches as illusions of binaural unmasking II: the Fourcin pitch and the Dichotic Repetition Pitch. *J Acoust Soc Am* 103:3525–3539.
- Culling JF, Summerfield Q, Marshall DH (1998b) Dichotic pitches as illusions of binaural unmasking I: Huggin’s pitch and the “Binaural Edge Pitch.” *J Acoust Soc Am* 103: 3509–3526.
- Dai H, Nguyen Q, Kidd GJ, Feth LL, Green DM (1996) Phase independence of pitch produced by narrow-band signals. *J Acoust Soc Am* 100:2349–2351.
- Dau T, Püschel D, Kohlrausch A (1996) A quantitative model of the “effective” signal processing in the auditory system. I. Model structure. *J Acoust Soc Am* 99:3615–3622.
- Davis H, Silverman SR, McAuliffe DR (1951) Some observations on pitch and frequency. *J Acoust Soc Am* 23:40–42.
- de Boer E (1956) On the “residue” in hearing. PhD Thesis.
- de Boer E (1976) On the “residue” and auditory pitch perception. In: Keidel WD, Neff WD (eds), *Handbook of Sensory Physiology*, Vol V-3. Berlin: Springer, pp. 479–583.
- de Boer E (1977) Pitch theories unified. In: Evans EF, and Wilson JP (eds), *Psychophysics and Physiology of Hearing*. London: Academic Press, pp. 323–334.



- de Cheveigné A (1989) Pitch and the narrowed autocoincidence histogram. *Proc ICMPC*, Kyoto, 67–70.
- de Cheveigné A (1993) Separation of concurrent harmonic sounds: fundamental frequency estimation and a time-domain cancellation model of auditory processing. *J Acoust Soc Am* 93:3271–3290.
- de Cheveigné A (1997a) Concurrent vowel identification III: A neural model of harmonic interference cancellation. *J Acoust Soc Am* 101:2857–2865.
- de Cheveigné A (1997b) Harmonic fusion and pitch shifts of inharmonic partials. *J Acoust Soc Am* 102:1083–1087.
- de Cheveigné A (1998) Cancellation model of pitch perception. *J Acoust Soc Am* 103:1261–1271.
- de Cheveigné A (1999) Pitch shifts of mistuned partials: a time-domain model. *J Acoust Soc Am* 106:887–897.
- de Cheveigné A (2000) A model of the perceptual asymmetry between peaks and troughs of frequency modulation. *J Acoust Soc Am* 107:2645–2656.
- de Cheveigné A (2001) Correlation Network model of auditory processing. In: *Proceedings of the Workshop on Consistent & Reliable Acoustic Cues for Sound Analysis*, Aalborg (Denmark).
- de Cheveigné A, Kawahara H (1999) Multiple period estimation and pitch perception model. *Speech Commun* 27:175–185.
- de Cheveigné A, Kawahara H (2002) YIN, a fundamental frequency estimator for speech and music. *J Acoust Soc Am* 111:1917–1930.
- Delgutte B (1984) Speech coding in the auditory nerve: II. Processing schemes for vowel-like sounds. *J Acoust Soc Am* 75:879–886.
- Delgutte B (1996) Physiological models for basic auditory percepts. In: Hawkins HL, McMullen TA, Popper AN, Fay RR (eds), *Auditory Computation*. New York: Springer, pp. 157–220.
- Demany L, Armand F (1984) The perceptual reality of tone chroma in early infancy. *J Acoust Soc Am* 76:57–66.
- Demany L, Clément S (1997) The perception of frequency peaks and troughs in wide frequency modulations. IV. Effects of modulation waveform. *J Acoust Soc Am* 102:2935–2944.
- Demany L, Ramos C (2004) Informational masking and pitch memory: perceiving a change in a non-perceived tone. *Proc CFA/DAGA*.
- Dooley GJ, Moore BCJ (1988) Detection of linear frequency glides as a function of frequency and duration. *J Acoust Soc Am* 84:2045–2057.
- Duchez M-E (1989) La notion musicale d'élément <<porteur de forme>>. Approche épistémologique et historique. In: McAdams S, Deliège I (eds), *La Musique et les Sciences Cognitives*. Liège: Pierre Mardaga, pp. 285–303.
- Duifhuis H, Willems LF, Sluyter RJ (1982) Measurement of pitch in speech: an implementation of Goldstein's theory of pitch perception. *J Acoust Soc Am* 71:1568–1580.
- Durlach NI (1963) Equalization and cancellation theory of binaural masking-level differences. *J Acoust Soc Am* 35:1206–1218.
- Du Verney JG (1683) *Traité de l'organe de l'ouïe, contenant la structure, les usages et les maladies de toutes les parties de l'oreille*. Paris.
- Elhilali M, Klein DJ, Fritz JB, Simon JZ, Shamma SA (2005) The enigma of cortical responses: slow yet precise. In: Pressnitzer D, de Cheveigné A, McAdams S, Collet L (eds), *Auditory Signal Processing: Psychophysics, physiology and modeling*. New York: Springer, pp. 485–494.

- Evans EF (1978) Place and time coding of frequency in the peripheral auditory system: some physiological pros and cons. *Audiology* 17:369–420.
- Evans EF (1986) Cochlear nerve fibre temporal discharge patterns, cochlear frequency selectivity and the dominant region for pitch. In: Moore BCJ, Patterson RD (eds), *Auditory Frequency Selectivity*. New York: Plenum Press, pp. 253–264.
- Fletcher H (1924) The physical criterion for determining the pitch of a musical tone. *Phys Rev* (reprinted in Shubert, 1979, 135–145) 23:427–437.
- Fourier JBJ (1822) *Traité analytique de la chaleur*. Paris: Didot.
- Gábor D (1947) Acoustical quanta and the theory of hearing. *Nature* 159:591–594.
- Galambos R, Davis H (1943) The response of single auditory-nerve fibers to acoustic stimulation. *J Neurophysiol* 6:39–57.
- Galilei G (1638) *Mathematical discourses concerning two new sciences relating to mechanics and local motion, in four dialogues*. Translated by Weston, London: Hooke (reprinted in Lindsay, 1973, pp. 40–61).
- Gerson A, Goldstein JL (1978) Evidence for a general template in central optimal processing for pitch of complex tones. *J Acoust Soc Am* 63:498–510.
- Gockel H, Moore BCJ, Carlyon RP (2001) Influence of rate of change of frequency on the overall pitch of frequency-modulated tones. *J Acoust Soc Am* 109:701–712.
- Goldstein JL (1970) Aural combination tones. In: Plomp R, Smoorenburg GF (eds), *Frequency Analysis and Periodicity Detection in Hearing*. Leiden: Sijthoff, pp. 230–247.
- Goldstein JL (1973) An optimum processor theory for the central formation of the pitch of complex tones. *J Acoust Soc Am* 54:1496–1516.
- Goldstein JL, Sruлович P (1977) Auditory-nerve spike intervals as an adequate basis for aural frequency measurement. In: Evans EF, Wilson JP (eds), *Psychophysics and Physiology of hearing*. London: Academic Press, pp. 337–347.
- Gray AA (1900) On a modification of the Helmholtz theory of hearing. *J Anat Physiol* 34:324–350.
- Grimault N, Micheyl C, Carlyon RP, Arthaud P, Collet L (2000) Influence of peripheral resolvability on the perceptual segregation of harmonic complex tones differing in fundamental frequency. *J Acoust Soc Am* 108:263–271.
- Große JH, Hall JW, III, Buss E (2002) Virtual pitch integration for asynchronous harmonics. *J Acoust Soc Am* 112:2956–2961.
- Hartmann WM (1993) On the origin of the enlarged melodic octave. *J Acoust Soc Am* 93:3400–3409.
- Hartmann WM (1996) Pitch, periodicity, and auditory organization. *J Acoust Soc Am* 100:3491–3502.
- Hartmann WM (1997) *Signals, sound and sensation*. Woodbury, NY: AIP.
- Hartmann WM, Doty SL (1996) On the pitches of the components of a complex tone. *J Acoust Soc Am* 99:567–578.
- Hartmann WM, Klein MA (1980) Theory of frequency modulation detection for low modulation frequencies. *J Acoust Soc Am* 67:935–946.
- Haykin S (1999) *Neural Networks, A Comprehensive Foundation*. Upper Saddle River, NJ: Prentice Hall.
- Hebb DO (1949) *The Organization of Behavior*. New York: John Wiley & Sons.
- Hebb DO (1959) A neuropsychological theory. In: Koch S (ed), *Psychology, A Study of a Science*, Vol. I. New York: McGraw-Hill, pp. 622–643.
- Heinz MG, Colburn HS, Carney LH (2001) Evaluating auditory performance limits: I. One-parameter discrimination using a computational model for the auditory nerve. *Neural Comput* 13:2273–2316. © 2001 by the Massachusetts Institute of Technology.

- Hess W (1983) Pitch determination of speech signals. Berlin: Springer.
- Hermes DJ (1988) Measurement of pitch by subharmonic summation. *J Acoust Soc Am* 83:257–264.
- Hewitt MJ, Meddis R (1994) A computer model of amplitude-modulation sensitivity of single units in the inferior colliculus. *J Acoust Soc Am* 95:2145–2159.
- Hewitt MJ, Meddis R, Shackleton TM (1992) A computer model of a cochlear nucleus stellate cell. Responses to amplitude-modulated and pure-tone stimuli. *J Acoust Soc Am* 91:2096–2109.
- Hirsh I (1948) The influence of interaural phase on interaural summation and inhibition. *J Acoust Soc Am* 20:536–544.
- Hounshell DA (1976) Bell and Gray: contrasts in style, politics and etiquette. *Proc IEEE* 64:1305–1314.
- Houtsma AJM, Goldstein JL (1972) The central origin of the pitch of complex tones. Evidence from musical interval recognition. *J Acoust Soc Am* 51:520–529.
- Houtsma AJM, Smurzynski J (1990) Pitch identification and discrimination for complex tones with many harmonics. *J Acoust Soc Am* 87:304–310.
- Huggins WH, Licklider JCR (1951) Place mechanisms of auditory frequency analysis. *J Acoust Soc Am* 23:290–299.
- Hunt FV (1992, original: 1978) Origins in acoustics. Woodbury, NY: Acoustical Society of America.
- Hurst CH (1895) A new theory of hearing. *Proc Trans Liverpool Biol Soc* 9:321–353 (and plate XX).
- Jeffress LA (1948) A place theory of sound localization. *J Comp Physiol Psychol* 41:35–39.
- Jenkins RA (1961) Perception of pitch, timbre and loudness. *J Acoust Soc Am* 33:1550–1557.
- Johnson DH (1980) The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. *J Acoust Soc Am* 68:1115–1122.
- Joris PX (2001) Sensitivity of inferior colliculus neurons to interaural time differences of broadband signals: comparison with auditory nerve firing. In: Breebaart DJ, Houtsma AJM, Kohlrausch A, Prijs VF, Schoonhoven R (eds), *Physiological and Psychophysical Bases of Auditory Function*. Maastricht: Shaker BV, pp. 177–183.
- Joris PX, Smith PH, Yin TCT (1998) Coincidence detection in the auditory system: 50 years after Jeffress. *Neuron* 21:1235–1238.
- Kaernbach C, Demany L (1998) Psychophysical evidence against the autocorrelation theory of pitch perception. *J Acoust Soc Am* 104:2298–2306.
- Köpl C (1997) Phase locking to high frequencies in the auditory nerve and cochlear nucleus magnocellularis of the barn owl *Tyto alba*. *J Neurosci* 17:3312–3321.
- Langner G (1981) Neuronal mechanisms for pitch analysis in the time domain. *Exp Brain Res* 44:450–454.
- Langner G (1998) Neuronal periodicity coding and pitch effects. In: Poon PWF, Brugge JF (eds), *Central Auditory Processing and Neural Modeling*. New York: Plenum.
- Langner G, Schreiner CE (1988) Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. *J. Neurophysiol* 60:1799–1822.
- Le Cat C-N (1758) *La Théorie de L'ouïe*: Supplément à cet Article du Traité des Sens. Paris: Vallat-la-Chapelle.
- Licklider JCR (1948) The influence of interaural phase relations upon the masking of speech by white noise. *J Acoust Soc Am* 20:150–159.
- Licklider JCR (1951) A duplex theory of pitch perception (reproduced in Schubert 1979, 155–160). *Experientia* 7:128–134.

- Licklider JCR (1959) Three auditory theories. In: Koch S (ed), *Psychology, A study of a Science*, Vol. I. New York: McGraw-Hill, pp. 41–144.
- Lindsay RB (1966) The story of acoustics. *J Acoust Soc Am* 39:629–644.
- Lindsay RB (1973) *Acoustics: historical and philosophical development*. Stroudsburg: Dowden, Hutchinson and Ross.
- Loeb GE, White MW, and Merzenich MM (1983) Spatial cross-correlation—a proposed mechanism for acoustic pitch perception. *Biol Cybern* 47:149–163.
- Lyon R (1984) Computational models of neural auditory processing. *Proc IEEE ICASSP*, 36.1(1–4).
- Maass W (1998) On the role of time and space in neural computation. *Lecture notes in computer science* 1450:72–83.
- Maass W, Natschläger T, Markram H (2003) Computation models for generic cortical microcircuits. In: Feng J (ed), *Computational Neuroscience: A Comprehensive Approach*. Boca Raton, FL: CRC Press, pp. 575–605.
- Macran HS (1902) *The harmonics of Aristoxenus*. Oxford: The Clarendon Press (re-printed 1990, Georg Olms Verlag, Hildesheim).
- Marozeau J, de Cheveigné A, McAdams S, and Winsberg S (2003) The dependency of timbre on fundamental frequency. *J Acoust Soc Am* 114:2946–2957.
- Martens JP (1984) Comment on “Algorithm for extraction of pitch and pitch salience from complex tonal signals” [*J Acoust Soc Am* 71, 679–688 (1982)]. *J Acoust Soc Am* 75:626–628.
- McAlpine D, Jiang D, Palmer A (2001) A neural code for low-frequency sound localization in mammals. *Nat Neurosci* 4:396–401.
- McKay CM, Carlyon RP (1999) Dual temporal pitch percepts from acoustic and electric amplitude-modulated pulse trains. *J Acoust Soc Am* 105:347–357.
- Meddis R (1988) Simulation of auditory-neural transduction: further studies. *J Acoust Soc Am* 83:1056–1063.
- Meddis R, Hewitt MJ (1991a) Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification. *J Acoust Soc Am* 89:2866–2882.
- Meddis R, Hewitt MJ (1991b) Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II: phase sensitivity. *J Acoust Soc Am* 89:2883–2894.
- Meddis R, Hewitt MJ (1992) Modeling the identification of concurrent vowels with different fundamental frequencies. *J Acoust Soc Am* 91:233–245.
- Meddis R, O’Mard L (1997) A unitary model of pitch perception. *J Acoust Soc Am* 102:1811–1820.
- Mersenne M (1636) *Harmonie Universelle*. Paris: Cramoisy (reprinted 1975, Paris: Editions du CNRS).
- Michéyl C, Carlyon RP (1998) Effects of temporal fringes on fundamental-frequency discrimination. *J Acoust Soc Am* 104:3006–3018.
- Miyazaki K (1990) The speed of musical pitch identification by absolute-pitch possessors. *Music Percept* 8:177–188.
- Moore BCJ (1973) Frequency difference limens for short-duration tones. *J Acoust Soc Am* 54:610–619.
- Moore BCJ (1977) *An Introduction to the Psychology of Hearing*. London: Academic Press (first edition).
- Moore BCJ (2003) *An introduction to the psychology of hearing*. London: Academic Press (fifth edition).
- Moore BCJ, Sek A (1994) Effects of carrier frequency and background noise on the detection of mixed modulation. *J Acoust Soc Am* 96:741–751.

- Nelken I, Ulanovsky N, Las L, Bar-Yosef O, Anderson M, Chechik G, Tishby N, Young E (2005) Transformation of stimulus representations in the ascending auditory system. In: Pressnitzer D, de Cheveigné A, McAdams S, Collet L (eds), *Auditory Signal Processing: Psychophysics, Physiology and Modeling*. New York: Springer, pp. 265–274.
- Newman EB, Stevens SS, and Davis H (1937) Factors in the production of aural harmonics and combination tones. *J Acoust Soc Am* 9:107–118.
- Noll AM (1967) Cepstrum pitch determination. *J Acoust Soc Am* 41:293–309.
- Nordmark J (1963) Some analogies between pitch and lateralization phenomena. *J Acoust Soc Am* 35:1544–1547.
- Nordmark JO (1968) Mechanisms of frequency discrimination. *J Acoust Soc Am* 44:1533–1540.
- Nordmark JO (1970) Time and frequency analysis. In: Tobias JV (ed), *Foundations of Modern Auditory Theory*. New York: Academic Press, pp. 55–83.
- Ohgushi K (1978) On the role of spatial and temporal cues in the perception of the pitch of complex tones. *J Acoust Soc Am* 64:764–771.
- Ohm GS (1843) On the definition of a tone with the associated theory of the siren and similar sound producing devices. *Poggendorf's Annalen der Physik und Chemie* 59: 497ff (translated and reprinted in Lindsay, 1973, pp. 242–247).
- Okada M, Kashino M (2003) The role of spectral change detectors in temporal order judgment of tones. *NeuroReport* 14:261–264.
- Oxenham A, Bernstein LR, Micheyl C (2005) Pitch perception of complex tones within and across ears and frequency regions. In: Pressnitzer D, de Cheveigné A, McAdams S, Collet L (eds), *Auditory Signal Processing: Physiology, Psychophysics and Modeling*. New York: Springer, pp. 126–135.
- Parncutt R (1988) Revision of Terhardt's psychoacoustical model of the roots of a musical chord. *Music Percept* 6:65–94.
- Parsons TW (1976) Separation of speech from interfering speech by means of harmonic selection. *J Acoust Soc Am* 60:911–918.
- Patterson RD (1987) A pulse ribbon model of monaural phase perception. *J Acoust Soc Am* 82:1560–1586.
- Patterson RD (1994a) The sound of a sinusoid: time-domain models. *J Acoust Soc Am* 96:1419–1428.
- Patterson RD (1994b) The sound of a sinusoid: spectral models. *J Acoust Soc Am* 96: 1409–1418.
- Patterson RD, Nimmo-Smith I (1986) Thinning periodicity detectors for modulated pulse streams. In: Moore BCJ, Patterson RD (eds), *Auditory Frequency Selectivity*. New York: Plenum Press, pp. 299–307.
- Patterson RD, Robinson K, Holdsworth J, McKeown D, Zhang C, Allerhand M (1992) Complex sounds and auditory images. In: Cazals Y, Horner K, Demany L (eds), *Auditory Physiology and Perception*. Oxford: Pergamon Press, pp. 429–446.
- Plack CJ, Carlyon RP (1995) Differences in frequency detection and fundamental frequency discrimination between complex tones consisting of resolved and unresolved harmonics. *J Acoust Soc Am* 98:1355–1364.
- Plack CJ, White LJ (2000a) Perceived continuity and pitch perception. *J Acoust Soc Am* 108:1162–1169.
- Plack CJ, White LJ (2000b) Pitch matches between unresolved complex tones differing by a single interpulse interval. *J Acoust Soc Am* 108:696–705.

- Plomp R (1964) The ear as a frequency analyzer. *J Acoust Soc Am* 36:1628–1636.
- Plomp R (1965) Detectability threshold for combination tones. *J Acoust Soc Am* 37: 1110–1123.
- Plomp R (1967a) Pitch of complex tones. *J Acoust Soc Am* 41:1526–1533.
- Plomp R (1967b) Beats of mistuned consonances. *J Acoust Soc Am* 42:462–474.
- Plomp R (1970) Timbre as a multidimensional attribute of complex tones. In: Plomp R, Smoorenburg GF (eds), *Frequency Analysis and Periodicity Detection in Hearing*. Leiden: Sijthoff, pp. 397–414.
- Plomp R (1976) *Aspects of tone sensation*. London: Academic Press.
- Plomp R, Levelt WJM (1965) Tonal consonance and critical bandwidth. *J Acoust Soc Am* 38:545–560.
- Pressnitzer D, Patterson RD (2001) Distortion products and the pitch of harmonic complex tones. In: Breebaart DJ, Houtsma AJM, Kohlrausch A, Prijs VF, Schoonhoven R (eds), *Physiological and Psychophysical Bases of Auditory Function*. Maastricht: Shaker, pp. 97–104.
- Pressnitzer D, Patterson RD, Krumbholz K (2001) The lower limit of melodic pitch. *J Acoust Soc Am* 109:2074–2084.
- Pressnitzer D, Winter IM, de Cheveigné A (2002) Perceptual pitch shift for sounds with similar waveform autocorrelation. *Acoust Res Lett Online* 3:1–6.
- Pressnitzer D, de Cheveigné A, Winter IM (2004) Physiological correlates of the perceptual pitch shift of sounds with similar waveform autocorrelation. *Acoust Res Lett Online* 5:1–6.
- Raatgever J, Bilsen FA (1986) A central spectrum model of binaural processing. Evidence from dichotic pitch. *J Acoust Soc Am* 80:429–441.
- Rameau J-P (1750) *Démonstration du principe de l'harmonie*, Paris: Durand [reproduced in E.R. Jacobi (1968) *Jean-Philippe Rameau, Complete theoretical writings*, V3, American Institute of Musicology, pp. 154–254].
- Rayleigh Lord (1896) *The theory of sound* (2nd ed., 1945 reissue). New York: Dover.
- Ritsma RJ (1967) Frequencies dominant in the perception of the pitch of complex tones. *J Acoust Soc Am* 42:191–198.
- Roederer JG (1975) *Introduction to the Physics and Psychophysics of Music*. New York: Springer.
- Rose JE, Brugge JF, Anderson DJ, Hind JE (1967) Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey. *J Neurophysiol* 30:769–793.
- Ross MJ, Shaffer HL, Cohen A, Freudberg R, Manley HJ (1974) Average magnitude difference function pitch extractor. *IEEE Trans ASSP* 22:353–362.
- Ruggero MA (1973) Response to noise of auditory nerve fibers in the squirrel monkey. *J Neurophysiol* 36:569–587.
- Ruggero MA (1992) Physiology of the auditory nerve. In: Popper AN, Fay RR (eds), *The Mammalian Auditory Pathway: Neurophysiology*. New York: Springer, pp. 34–93.
- Rutherford E (1886) A new theory of hearing. *J Anat Physiol* 21:166–168.
- Sabine WC (1907) Melody and the origin of the musical scale. In: Hunt FV (ed), *Collected Papers on Acoustics by Wallace Clement Sabine* (1964). New York: Dover, pp. 107–116.
- Sachs MB, Young ED (1979) Encoding of steady-state vowels in the auditory nerve: representation in terms of discharge rate. *J Acoust Soc Am* 66:470–479.
- Sahey TL, Nodar RH, Musiek FE (1997) *Efferent Auditory System*. San Diego: Singular.

- Sauveur J (1701) *Système général des intervalles du son*, Mémoires de l'Académie Royale des Sciences 279–300:347–354 (translated and reprinted in Lindsay, 1973, pp. 88–94).
- Scheffers MTM (1983) *Sifting vowels*. PhD Thesis, University of Gröningen.
- Schouten JF (1938) The perception of subjective tones. *Proc Kon Acad Wetensch (Neth.)* 41:1086–1094 (reprinted in Schubert 1979, 146–154).
- Schouten JF (1940a) The residue, a new component in subjective sound analysis. *Proc Kon Acad Wetensch (Neth.)* 43:356–356.
- Schouten JF (1940b) The residue and the mechanism of hearing. *Proc Kon Acad Wetensch (Neth.)* 43:991–999.
- Schouten JF (1940c) The perception of pitch. *Philips Tech Rev* 5:286–294.
- Schouten JF (1970) The residue revisited. In: Plomp R, Smoorenburg GF (eds), *Frequency Analysis and Periodicity Detection in Hearing*. London: Sijthoff, pp. 41–58.
- Schouten JF, Ritsma RJ, Cardozo BL (1962) Pitch of the residue. *J Acoust Soc Am* 34:1418–1424.
- Schroeder MR (1968) Period histogram and product spectrum: new methods for fundamental-frequency measurement. *J Acoust Soc Am* 43:829–834.
- Schubert ED (1978) History of research on hearing. In: Carterette EC, Friedman MP (eds), *Handbook of Perception*, Vol. IV. New York: Academic Press, pp. 41–80.
- Schubert ED (1979) *Psychological acoustics (Benchmark papers in Acoustics, Vol 13)*. Stroudsburg, PA: Dowden, Hutchinson & Ross.
- Sek A, Moore BCJ (1999) Discrimination of frequency steps linked by glides of various durations. *J Acoust Soc Am* 106:351–359.
- Semal C, Demany L (1990) The upper limit of musical pitch. *Music Percept* 8:165–176.
- Shackleton TM, Carlyon RP (1994) The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. *J Acoust Soc Am* 95:3529–3540.
- Shamma SA (1985) Speech processing in the auditory system II: Lateral inhibition and the central processing of speech evoked activity in the auditory nerve. *J Acoust Soc Am* 78:1622–1632.
- Shamma S, Klein D (2000) The case of the missing pitch templates: how harmonic templates emerge in the early auditory system. *J Acoust Soc Am* 107:2631–2644.
- Shamma SA, Shen N, Gopalaswamy P (1989) Stereausis: binaural processing without neural delays. *J Acoust Soc Am* 86:989–1006.
- Shera CA, Guinan JJ, Oxenham AJ (2002) Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. *Proc Natl Acad Sci USA* 99:3318–3323.
- Siebert WM (1968) Stimulus transformations in the auditory system. In: Kolars PA, Eden M (eds), *Recognizing Patterns*. Cambridge, MA: MIT Press, pp. 104–133.
- Siebert WM (1970) Frequency discrimination in the auditory system: place or periodicity mechanisms. *Proc IEEE* 58:723–730.
- Slaney M (1990) A perceptual pitch detector. *Proc ICASSP*, 357–360.
- Smoorenburg GF (1970) Pitch perception of two-frequency stimuli. *J Acoust Soc Am* 48:924–942.
- Srulovicz P, Goldstein JL (1983) A central spectrum model: a synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum. *J Acoust Soc Am* 73:1266–1276.
- Tannery M-P, de Waard C (1970) *Correspondance du P. Marin Mersenne*, Vol. XI (1642). Paris: Editions du CNRS.

- Tasaki I (1954) Nerve impulses in individual auditory nerve fibers of guinea pig. *J Neurophysiol* 17:97–122.
- Terhardt E (1974) Pitch, consonance and harmony. *J Acoust Soc Am* 55:1061–1069.
- Terhardt E (1978) Psychoacoustic evaluation of musical sounds. *Percept Psychophys* 23: 483–492.
- Terhardt E (1979) Calculating virtual pitch. *Hear Res* 1:155–182.
- Terhardt E (1991) Music perception and sensory information acquisition: relationships and low-level analogies. *Music Percept* 8:217–240.
- Terhardt E, Stoll G, Seewann M (1982) Algorithm for extraction of pitch and pitch salience from complex tonal signals. *J Acoust Soc Am* 71:679–688.
- Thompson SP (1882) On the function of the two ears in the perception of space. *Phil Mag* (S5) 13:406–416.
- Thorpe S, Fize F, Marlot C (1996) Speed of processing in the human visual system. *Nature* 381:520–522.
- Thurlow WR (1963) Perception of low auditory pitch: a multicue mediation theory. *Psychol Rev* 70:461–470.
- Tong YC, Blamey PJ, Dowell RC, Clark GM (1983) Psychophysical studies evaluating the feasibility of speech processing strategy for a multichannel cochlear implant. *J Acoust Soc Am* 74:73–80.
- Troland LT (1930) Psychophysiological considerations related to the theory of hearing. *J Acoust Soc Am* 1:301–310.
- Turner RS (1977) The Ohm-Seebeck dispute, Hermann von Helmholtz, and the origins of physiological acoustics. *Brit J Hist Sci* 10:1–24.
- van Noorden L (1982) Two channel pitch perception. In Clynes M (ed), *Music, Mind, and Brain*. London: Plenum Press, pp. 251–269.
- Versnel H, Shamma S (1998) Spectral-ripple representation of steady-state vowels. *J Acoust Soc Am* 103:5502–2514.
- von Békésy G, Rosenblith WA (1948) The early history of hearing—observations and theories. *J Acoust Soc Am* 20:727–748.
- von Helmholtz H (1857, translated by A.J. Ellis, reprinted in Warren & Warren 1968) *On the Physiological Causes of Harmony in Music*, pp. 25–60.
- von Helmholtz H (1877) *On the Sensations of Tone* (English translation A.J. Ellis, 1885, 1954). New York: Dover.
- Ward WD (1999) Absolute pitch. In: Deutsch D (ed), *The Psychology of Music*. Orlando: Academic Press, pp. 265–298.
- Warren RM, Warren RP (1968) *Helmholtz on Perception: Its Physiology and Development*. New York: John Wiley & Sons.
- Warren JD, Uppenkamp S, Patterson RD, Griffith TD (2003) Separating pitch chroma and pitch height in the human brain. *Proc Natl Acad Sci USA* 100:10038–19942.
- Watt HJ (1917) *The Psychology of Sound*. Cambridge: Cambridge University Press.
- Wegel RL, Lane CE (1924) The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear. *Physical Rev* 23:266–285 (reproduced in Schubert 1979, 201–211).
- Weintraub M (1985) A theory and computational model of auditory monaural sound separation. PhD Thesis, Stanford University.
- Wever EG (1949) *Theory of Hearing*. New York: Dover.
- Wever EG, Bray CW (1930) The nature of acoustic response: the relation between sound frequency and frequency of impulses in the auditory nerve. *J Exp Psychol* 13:373–387.



- Whitfield IC (1970) Central nervous processing in relation to spatio-temporal discrimination of auditory patterns. In: Plomp R, Smoorenburg GF (eds), *Frequency Analysis and Periodicity Detection in Hearing*. Leiden: Sijthoff, pp. 136–152.
- Wiegrefe L (2001) Searching for the time constant of neural pitch integration. *J Acoust Soc Am* 109:1082–1091.
- Wiegrefe L, Meddis R (2004) The representation of periodic sounds in simulated sustained chopper units of the ventral cochlear nucleus. *J Acoust Soc Am* 116:1207–1218.
- Wiegrefe L, Patterson RD, Demany L, Carlyon RP (1998) Temporal dynamics of pitch strength in regular interval noises. *J Acoust Soc Am* 104:2307–2313.
- Wiegrefe L, Stein A, Meddis R (2005) Coding of pitch and amplitude modulation in the auditory brainstem: one common mechanism? In: Pressnitzer D, de Cheveigné A, McAdams S, Collet L (eds), *Auditory Signal Processing: Psychophysics, Physiology and Modeling*. New York: Springer, pp. 117–125.
- Wightman FL (1973) The pattern-transformation model of pitch. *J Acoust Soc Am* 54: 407–416.
- Yost WA (1996) Pitch strength of iterated rippled noise. *J Acoust Soc Am* 100:3329–3335.
- Young T (1800) Outlines of experiments and inquiries respecting sound and light. *Philos Trans of the Royal Society of London* 90:106–150 (and plates).
- Young ED, Sachs MB (1979) Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. *J Acoust Soc Am* 66:1381–1403.
- Zwicker E (1970) Masking and psychoacoustical excitation as consequences of the ear's frequency analysis. In: Plomp R, Smoorenburg GF (eds), *Frequency Analysis and Periodicity Detection in Hearing*. Leiden: Sijthoff, pp. 376–396.

# Perception of Pitch by People with Cochlear Hearing Loss and by Cochlear Implant Users

BRIAN C.J. MOORE AND ROBERT P. CARLYON

## 1. Introduction

This chapter is concerned with the perception of pitch by people with cochlear hearing loss and by people with cochlear implants. These topics are of interest not only because of their clinical relevance, but also because they help us to understand the basic mechanisms of normal pitch perception. For both hearing-impaired people and cochlear implant users, we start with some basic considerations of how the representation of sounds in the auditory system differs from that in the normal auditory system. Experimental data are interpreted in the light of these differences.

## 2. Physiological Consequences of Cochlear Hearing Loss

Cochlear hearing loss results in a variety of changes in the way that sounds are represented in the auditory system. Four such changes are especially relevant for the perception of pitch:

1. Frequency selectivity is reduced; auditory filters are broader than normal (Pick et al. 1977; Glasberg and Moore 1986; Moore 1998). Hence, the excitation pattern evoked by a sinusoid is also broader than normal. According to the place theory (see Plack and Oxenham, Chapter 2; Winter, Chapter 4), this should lead to impaired frequency discrimination of sinusoids. Reduced frequency selectivity also presumably leads to a reduced ability to resolve partials in complex tones (although this has not been directly measured, to our knowledge), and this might adversely affect the perception of the pitch of complex tones; see Plack and Oxenham, Chapter 2; de Cheveigné, Chapter 6).
2. The precision of phase locking can be reduced (Woolf et al. 1981; Miller et al. 1999), although this has not always been found. According to the temporal theory (see Plack and Oxenham, Chapter 2; Winter, Chapter 4), reduced precision of phase locking should adversely affect frequency discrimination.

3. The propagation time of the traveling wave along the basilar membrane and the relative phase of the response at different places may differ from normal, because of loss of the “active mechanism,” structural abnormalities, or both (Ruggero 1994; Ruggero et al. 1996). This could adversely affect mechanisms for pitch perception based on cross-correlation of the outputs of different points on the basilar membrane (Loeb et al. 1983; Shamma 1985; Shamma and Klein 2000).
4. There may be regions within the cochlea where the inner hair cells (IHCs) and/or neurons are completely nonfunctional. These are referred to as “dead regions.” A dead region can be defined in terms of the characteristic frequencies (CFs) of the functioning IHCs and neurons adjacent to the dead region. When a tone has a frequency falling within a dead region, it may be detected via a remote region. The peak in the neural excitation pattern may occur at a place very different from that normally associated with that frequency. The place theory predicts that the perceived pitch of the tone in such a case should be very different from normal.

### 3. Frequency Discrimination of Pure Tones

#### *3.1. Basic Mechanisms of Frequency Discrimination*

The basic mechanisms underlying the frequency discrimination of pure tones have been reviewed earlier in this book—see Plack and Oxenham, Chapter 2. However, for completeness a brief summary is given here. It has been proposed that the frequency discrimination of steady pulsed tones by normally hearing listeners is largely based on temporal information (cues derived from phase locking) for frequencies up to 4 to 5 kHz (Moore 1973a,b, 1974, 2003; Goldstein and Srulovicz 1977; Sek and Moore 1995; Micheyl et al. 1998; Heinz et al. 2001). Above 4 to 5 kHz, frequency discrimination is thought to depend mainly on place mechanisms, based on changes in the excitation pattern (Moore 1973b; Sek and Moore 1995), although residual phase locking may play some role (Heinz et al. 2001). The mechanisms underlying the detection of frequency modulation (FM) of sinusoidal carriers are thought to depend on the modulation rate. For sinusoidal modulation with rates above about 10 Hz, detection is probably largely based on excitation-pattern cues (Zwicker 1956; Zwicker and Fastl 1990; Moore and Sek 1994, 1995; Saberi and Hafter 1995; Sek and Moore 1995). FM results in modulation of the excitation level at each place on the pattern, so the FM is effectively transformed into amplitude modulation (AM). Thus, the FM can be detected as AM, either by using information from the single point on the excitation pattern where the AM is greatest (Zwicker 1956; Zwicker and Fastl 1990) or by combining information from different parts of the excitation pattern (Moore and Sek 1994).

For very low FM rates (around 2 Hz), temporal information may also play a role (Moore and Sek 1995, 1996; Plack and Carlyon 1995; Sek and Moore

1995); the short-term pattern of phase locking can be used to estimate the momentary frequency, and changes in phase locking over time indicate that FM is present. A similar temporal mechanism probably plays a role in the detection of FM of the fundamental frequency (F0) of harmonic complex tones, when those tones are bandpass filtered so as to contain only unresolved harmonics (Plack and Carlyon 1994, 1995; Shackleton and Carlyon 1994; Carlyon et al. 2000). Indeed, for such tones, place information is not available at all, so subjects are forced to rely on temporal information.

The temporal mechanism may become less effective for modulation rates above about 5 Hz because it is “sluggish,” and cannot follow rapid changes in frequency. Consistent with this idea, thresholds for detecting FM of the F0 of harmonic complex tones containing only unresolved harmonics increase with increasing modulation rate over the range 1 to 20 Hz, reaching 20% (defined as the peak deviation in F0 divided by the mean F0) for a modulation rate of 20 Hz (Carlyon et al. 2000). In the case of sinusoidal carriers, performance does not change much with increasing modulation rate (Zwicker and Fastl 1990; Moore and Sek 1995, 1996; Sek and Moore 1995), presumably because the place mechanism “takes over” from the temporal mechanism for modulation rates above 5 to 10 Hz.

### *3.2. Frequency Difference Limens (FDLs) Measured Using Subjects with Cochlear Hearing Loss*

The frequency difference limen (FDL) is a measure of the ability to discriminate the frequency of steady pure tones, presented successively. Many studies have measured FDLs in people with cochlear hearing loss (Gengel 1973; Tyler et al. 1983; Hall and Wood 1984; Freyman and Nelson 1986, 1987, 1991; Moore and Glasberg 1986; Moore and Peters 1992; Simon and Yund 1993). The results have generally shown that frequency discrimination is adversely affected by cochlear hearing loss. However, there is considerable variability across individuals and the size of the FDL is not strongly correlated with the absolute threshold at the test frequency. Simon and Yund (1993) measured FDLs separately for each ear of subjects with bilateral cochlear damage and found that FDLs could be markedly different for the two ears at frequencies for which absolute thresholds were the same. They also found that FDLs could be the same for the two ears when absolute thresholds were different.

Tyler et al. (1983) compared FDLs and frequency selectivity measured using psychophysical tuning curves (PTCs). They found a low correlation between the two. They concluded that frequency discrimination was not closely related to frequency selectivity, suggesting that place models were not adequate to explain the data. Moore and Peters (1992) measured FDLs for four groups of subjects: young normally hearing, young hearing impaired, elderly with near-normal hearing, and elderly hearing impaired. The auditory filter shapes of the subjects had been estimated in earlier experiments using the notched-noise method (Glasberg and Moore 1990), for center frequencies ( $f_c$ ) of 100,

200, 400 and 800 Hz. The FDLs for both impaired groups were higher than for the young normal group at all  $f_c$ s (50 to 4000 Hz). The FDLs for the elderly group with near-normal hearing were intermediate. The FDLs at a given center frequency were generally only weakly correlated with the sharpness of the auditory filter at that center frequency, and some subjects with broad filters at low frequencies had near-normal FDLs at low frequencies. These results suggest a partial dissociation of frequency selectivity and frequency discrimination of pure tones.

Overall, the results of these experiments do not provide strong support for place models of frequency discrimination. This is consistent with the conclusion presented earlier, that FDLs for normally hearing people are determined mainly by temporal mechanisms for frequencies up to about 5 kHz. An alternative way of accounting for the fact that cochlear hearing loss results in larger-than-normal FDLs is in terms of loss of neural synchrony (phase locking) in the auditory nerve. Goldstein and Srulovicz (1977) described a model for frequency discrimination based on the use of information from the interspike intervals in the auditory nerve. This model was able to account for the way that FDLs depend on frequency and duration for normally hearing subjects. Wakefield and Nelson (1985) showed that a simple extension to this model, taking into account the fact that phase locking gets slightly more precise as sound level increases, allowed the model to predict the effects of level on FDLs. They also applied the model to FDLs measured as a function of level in subjects with high-frequency hearing loss, presumably resulting from cochlear damage. They were able to predict the results of the hearing-impaired subjects by assuming that neural synchrony was reduced in neurons with characteristic frequencies corresponding to the region of hearing loss. Of course, this does not prove that loss of synchrony is the cause of the larger FDLs, but it does demonstrate that loss of synchrony is a plausible candidate.

Yet another possibility is that the central mechanisms involved in the analysis of phase-locking information make use of differences in the preferred time of firing of neurons with different characteristic frequencies; these time differences arise from the propagation time of the traveling wave on the basilar membrane (Loeb et al. 1983; Shamma 1985). The propagation time along the basilar membrane can be affected by cochlear damage (Ruggero 1994; Ruggero et al. 1996), and this could disrupt the processing of the temporal information by central mechanisms.

### *3.3. Frequency Modulation Detection Limens Measured Using Hearing-Impaired Subjects*

The frequency modulation detection limen (FMDL) is a measure of the ability to detect frequency modulation. Usually, a two-interval forced-choice task is used; one interval contains an unmodulated tone, and the other contains a frequency modulated tone and the task of the subject is to identify the interval with the modulated tone. Zurek and Formby (1981) measured FMDLs in 10 subjects

with sensorineural hearing loss (assumed to be mainly of cochlear origin) using a 3-Hz modulation rate and frequencies between 125 and 4000 Hz. Subjects were tested at a sensation level (SL) of 25 dB, a level above which performance was found (in pilot studies) to be roughly independent of level. The FMDLs tended to increase with increasing hearing loss at a given frequency. For a given degree of hearing loss, the worsening of performance with increasing hearing loss was greater at low frequencies than at high frequencies.

Zurek and Formby suggested two possible explanations for the greater effect at low frequencies. The first is based on the assumption that two mechanisms are involved in coding frequency, a temporal mechanism at low frequencies and a place mechanism at high frequencies. The temporal mechanism may be more disrupted by hearing loss than the place mechanism. An alternative possibility is that absolute thresholds at low frequencies do not provide an accurate indicator of the extent of cochlear damage, since these thresholds may be mediated by neurons with CFs above the test frequency. In extreme cases, there may be a dead region at low frequencies (Thornton and Abbas 1980; Florentine and Houtsma 1983; Turner et al. 1983; Moore et al. 2000; Moore 2001). When a sinusoid has a frequency that falls within a dead region, it appears to evoke a less clear pitch than normal, and sometimes does not even sound like a tone (Huss and Moore 2005a,b) (see Section 4 for details).

Moore and Glasberg (1986) measured both FMDLs and thresholds for detecting amplitude modulation, using a 4-Hz modulation rate. Subjects with moderate unilateral and bilateral cochlear impairments were tested. Stimuli were presented at a fixed level of 80 dB SPL, which was at least 10 dB above the absolute threshold. The FMDLs were larger for the impaired than for the normal ears, by an average factor of 3.8 for a frequency of 500 Hz and 1.5 for a frequency of 2000 Hz, although the average hearing loss was similar for these two frequencies. The greater effect at low frequencies is consistent with the results of Zurek and Formby, described earlier. The amplitude-modulation detection thresholds were not very different for the normal and impaired ears. These thresholds provide an estimate of the smallest detectable change in excitation level. Moore and Glasberg also used the notched-noise method (Patterson 1976; Glasberg and Moore 1990) to estimate the slopes of the auditory filters, at each test frequency. The slopes, together with the amplitude-modulation detection thresholds, were used to predict the FMDLs on the basis of Zwicker's excitation-pattern model. The obtained FMDLs were reasonably close to the predicted values. In other words, the results were consistent with the excitation-pattern model.

Grant (1987) measured FMDLs for three normally hearing subjects and three subjects with profound hearing losses. The sinusoidal carrier was modulated in frequency by a triangle function three times per second. Stimuli were presented at 30 dB SL for the normal subjects and at a "comfortable listening level" (110 to 135 dB SPL) for the impaired subjects. For all carrier frequencies (100 to 1000 Hz), FMDLs were larger, by an average factor of 9.5, for the hearing-impaired subjects than for the normally hearing subjects. Grant also measured

FMDLs when the stimuli were simultaneously amplitude modulated by a noise that was lowpass filtered at 3 Hz. The slow random amplitude fluctuations produced by this amplitude modulation would be expected to impair the use of cues for frequency modulation detection based on changes in excitation level. Consistent with the predictions of the excitation-pattern model, the random amplitude modulation led to increased FMDLs. Interestingly, the increase was much greater for the hearing-impaired than for the normally hearing subjects. When the random amplitude modulation was present, thresholds for the hearing-impaired subjects were about 16 times those for the normally hearing subjects. This large effect may have been partly due to the loss of cochlear compression in the hearing-impaired subjects (Moore 1998). This would magnify the effective “internal” modulation depth produced by the AM (Moore et al. 1996).

It is likely that, for low modulation rates, normally hearing subjects can extract information about frequency modulation both from changes in excitation level and from phase locking (Moore and Sek 1995; Sek and Moore 1995). The random amplitude modulation disrupts the use of changes in excitation level, but does not markedly affect the use of phase-locking cues. The profoundly hearing-impaired subjects of Grant appear to have been relying mainly or exclusively on changes in excitation level. Hence, the random amplitude modulation had severe adverse effects on the FMDLs.

Lacher-Fougère and Demany (1998) measured FMDLs for a 500-Hz carrier, using modulation rates of 2 and 10 Hz. They tested five normally hearing subjects and seven subjects with cochlear hearing loss ranging from 30 dB to 75 dB at 500 Hz. Stimuli were presented at a “comfortable” loudness level. The subjects with losses up to 45 dB had thresholds that were about a factor of two larger than for the normally hearing subjects. The subjects with larger losses had thresholds up to ten times larger than normal. The effect of the hearing loss was similar for the two modulation rates. Lacher-Fougère and Demany suggested that cochlear hearing loss disrupts excitation-pattern (place) cues and phase-locking cues to a roughly equal extent.

Moore and Skrodzka (2002) measured FMDLs for three young subjects with normal hearing and four elderly subjects with cochlear hearing loss. Carrier frequencies were 0.25, 0.5, 1, 2, 4, and 6 kHz and modulation rates were 2, 5, 10, and 20 Hz. FM detection thresholds were measured both in the absence of AM, and with AM of a fixed depth ( $m = 0.33$ , corresponding to a peak-to-valley ratio of 6 dB) added in both intervals of a forced-choice trial. The added AM was intended to disrupt cues based on FM-induced AM in the excitation pattern. The results averaged across subjects are shown in Figure 7.1. Generally, the hearing-impaired subjects (filled symbols) performed markedly more poorly than the normally hearing subjects (open symbols). For the normally hearing subjects, the disruptive effect of the AM (triangles versus circles) tended to increase with increasing modulation rate, for carrier frequencies below 6 kHz, as found previously by Moore and Sek (1996). For the hearing-impaired subjects, the disruptive effect of the AM was generally larger than for the normally hearing subjects, and the magnitude of the disruption did not consistently

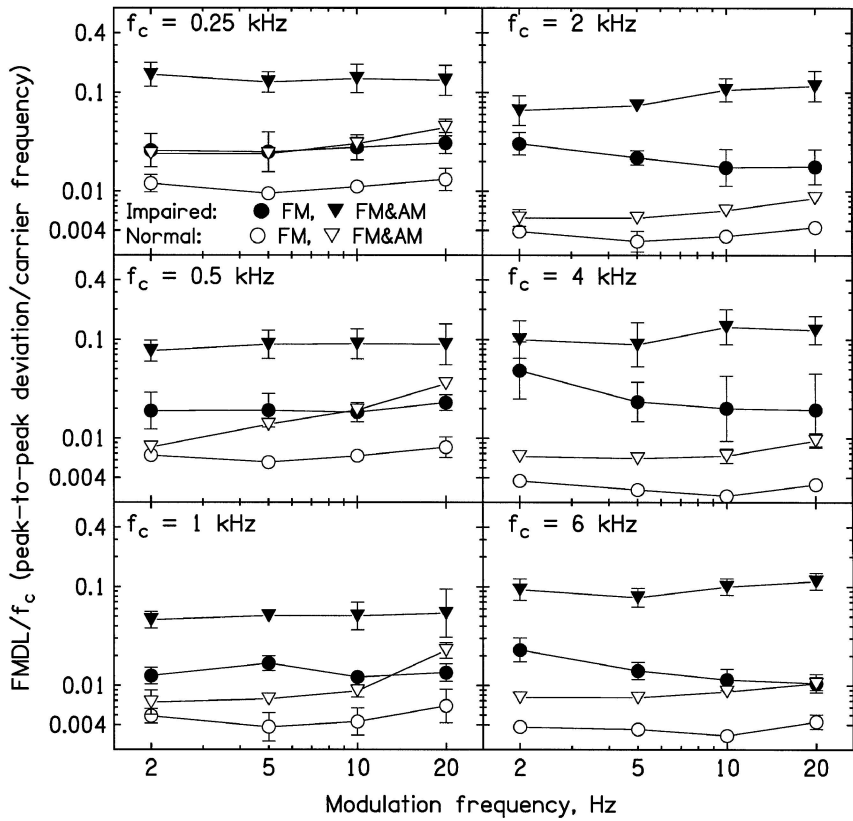


FIGURE 7.1. FMDLs plotted as a function of modulation frequency. Each panel shows results for one carrier frequency. Mean results are shown for normally hearing subjects (*open symbols*) and hearing-impaired subjects (*filled symbols*). FMDLs are shown without added AM (*circles*) and with added AM (*triangles*). Error bars indicate  $\pm$  one standard deviation across subjects. They are omitted when they would span a range less than 0.1 log units (corresponding to a ratio of 1.26).

increase with increasing modulation rate. For the 2-Hz modulation rate, the FMDL for the hearing-impaired subjects, averaged across the four lowest carrier frequencies, was a factor of 2.5 larger when AM was present than when it was absent. In contrast, the corresponding ratio for the normally hearing subjects was only 1.45. It has been argued in the past that the relatively small disruptive effect of AM at low modulation rates for normally hearing subjects reflects the use of temporal information (Moore and Sek 1996). The larger effect found for the hearing-impaired subjects suggests that they were not using temporal information effectively. Rather, the FMDLs were probably based largely on excitation-pattern cues (FM-induced AM in the excitation pattern), and these cues were strongly disrupted by the added AM. Overall, the results suggest that



cochlear hearing impairment adversely affects both temporal and excitation pattern mechanisms of FM detection.

In conclusion, FMDLs for hearing-impaired people are generally larger than normal. The larger thresholds may reflect both the broadening of the excitation pattern (reduced frequency selectivity) and disruption of cues based on phase locking.

#### 4. The Perception of Pure-Tone Pitch for Frequencies Falling in a Dead Region

In some people with cochlear damage restricted to low-frequency regions of the cochlea, it appears that there are no functioning IHCs and/or neurons with CFs corresponding to the frequency region of the loss. In other words, there may be a low-frequency dead region, as described earlier. In such cases, the detection of low-frequency tones is mediated by neurons with high CFs. One way of demonstrating this is by the measurement of PTCs. To measure a PTC, the signal is fixed in frequency and in level, usually at a level just above the absolute threshold, say, 10 dB SL. The masker can be either a sinusoid or a narrow band of noise. For each of several masker center frequencies, the level of the masker needed just to mask the signal is determined. Usually, the tip of the PTC (i.e., the frequency at which the masker level is lowest) lies close to the signal frequency. However, if the signal falls within a low-frequency dead region, the tip of the tuning curve lies well above the signal frequency. In other words, a masker centered well above the signal in frequency is more effective than a masker centered close to the signal frequency (Thornton and Abbas 1980; Florentine and Houtsma 1983; Turner et al. 1983; Moore et al. 2000; Moore and Alcántara 2001). For example, Florentine and Houtsma (1983) studied a subject with a moderate to severe unilateral low-frequency loss of cochlear origin. For a 1-kHz signal, the tip of the PTC fell between 2.2 and 2.85 kHz, depending on the exact level of the signal.

The perception of pitch in such subjects is of considerable theoretical interest. When there is a low-frequency dead region, a low-frequency pure tone cannot produce maximum neural excitation at the CF corresponding to its frequency, since there are no neural responses at that CF. The peak in the neural excitation pattern must occur at CFs higher than the frequency of the signal. If the place theory is correct, this should lead to marked upward shifts in the pitch of the tone. In fact, any tone whose frequency fell within a dead region should evoke the same pitch; this pitch should correspond to the CF immediately adjacent to the dead region. In most such cases studied, upward shifts in pitch were not observed and the pitches of tones whose frequencies fell in the dead region did shift with frequency in an appropriate way.

Florentine and Houtsma (1983) obtained pitch matches between the two ears of their unilaterally impaired subject. They presented the stimuli at levels just above absolute threshold, to minimize the spread of excitation along the basilar

membrane. Pitch shifts between the two ears were small. However, the variability of the pitch matches was rather large, indicating that the pitch in the impaired ear was not clear. Turner et al. (1983) studied six subjects with low-frequency cochlear losses. Three of their subjects showed PTCs with tips close to the signal frequency; they presumably had functioning IHCs with characteristic frequencies close to the signal frequency. The other three subjects showed PTCs with tips well above the signal frequency; they presumably had low-frequency dead regions. Pitch perception was studied either by pitch matching between the two ears (for subjects with unilateral losses) or by octave matching (for subjects with bilateral losses, but with some musical ability). The subjects whose PTCs had tips above the signal frequency gave results similar to those of the subjects whose PTCs had tips close to the signal frequency; no distinct pitch anomalies were observed.

A similar study was conducted by Huss et al. (2001) and Huss and Moore (2005b). Two tasks were used: a pitch-matching task and an octave-matching task. For the pitch-matching task, subjects were asked to match the perceived pitch of a pure tone with that of another fixed-frequency pure tone. The two tones were presented alternately. Matches were made across ears, to obtain a measure of diplacusis, and within one ear, to estimate the reliability of matching. For the octave-matching task, subjects were asked to adjust a tone of variable frequency so that it sounded one octave higher or lower than a fixed reference tone. Only a few subjects were able to perform this task reliably. The level for each frequency was chosen using a loudness model (Moore and Glasberg 1997), so as to give a fixed calculated loudness.

Results of the pitch-matching task for a subject with severe hearing loss in the right ear and a moderate high-frequency loss in the left ear are shown in Figure 7.2. On the basis of the test using “threshold equalizing noise” (TEN) described by Moore et al. (2000), and on the basis of measurement of PTCs (Moore and Alcántara 2001), this subject was diagnosed as having extensive low-frequency and high-frequency dead regions in the right ear, with an “island” of functioning IHCs around 3.5 kHz. The left ear had a dead region above about 4 kHz. Each x denotes one match, and means are shown by open circles. Matches within his better ear (top) were reasonably accurate at low frequencies, but became less accurate at high frequencies. Matches within his worse ear (middle), were more erratic, indicating a less clear pitch percept. Matches across ears, with the fixed tone in his worse ear (bottom), showed considerable variability, but also some consistent deviations. A fixed tone of 0.5 kHz in the worse ear was matched with a tone of about 3.5 kHz in the better ear. Generally, the matched frequency lay above the fixed frequency, for all fixed frequencies up to about 4 kHz, indicating upward pitch shifts in the worse ear.

The results of Florentine and Houtsma (1983) and of Turner et al. (1983) are hard to explain in terms of the traditional place theory. They show that a pure tone can evoke a low pitch even when there are no functioning IHCs or neurons with characteristic frequencies corresponding to that pitch. Their results are more readily explained in terms of the temporal theory; the pitch of the low-

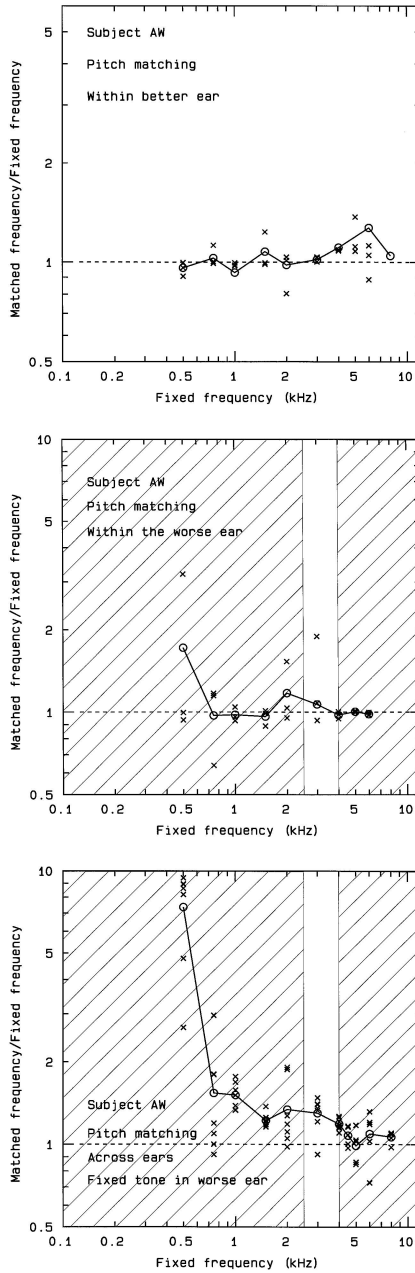


FIGURE 7.2. Results of the pitch-matching task for subject AW, who had extensive dead regions in his worse ear, shown by the *shaded areas*, and a moderate high-frequency loss without any dead region in his better ear. Each *x* denotes one match, and means are shown by *open circles*. Matches were made within his better ear (*top*), within his worse ear (*middle*), and across ears (*bottom*).

frequency tone may be coded in the temporal pattern of neural responses in neurons with characteristic frequencies above the signal frequency. However, the results of Huss et al. (2001) and Huss and Moore (2005b) show that such a low pitch is not always perceived. Perhaps the “correct” pitch is not perceived when there is too large a mismatch between temporal information and place information (see also Sections 9 and 10.1).

There have been a few studies of pitch perception in people with hearing losses that increase abruptly at high frequencies, who probably had dead regions at high frequencies. These subjects often report that high-frequency sinusoids do not have a distinct pitch, but sound like noises or buzzes (Villchur 1973; Moore et al. 1985b; Murray and Byrne 1986). However, Huss et al. (2001) and Huss and Moore (2005a) found that a tone with frequency corresponding to a dead region was not always described as sounding noise-like, and subjects without dead regions sometimes reported pure tones to sound noise-like. Subjective reports that pure tones sound noise-like may be taken as a hint that a dead region is present, but ratings of the clarity of the tonal percept cannot be used as a reliable indicator of dead regions.

McDermott et al. (1998) measured FDLs as a function of center frequency for subjects with near-normal hearing at low frequencies, but profound hearing loss at high frequencies. These subjects were assumed to have high-frequency dead regions. The mean levels of the stimuli were chosen to fall along an equal-loudness contour, and the level of each tone was roved over a range of  $\pm 3$  dB, to reduce the possibility of subjects using loudness changes as a cue. For tones falling in the frequency range of near-normal hearing, the DLs were typically about 2–3% of the center frequency. For tones falling well within the dead region, the DLs increased to about 10%. Interestingly, the DLs sometimes decreased slightly for frequencies just below the presumed boundary of the dead region, which McDermott et al. suggested might be the result of cortical over-representation of CFs just below the dead region. A similar effect has been reported by Thai-Van et al. (2003) for subjects with severe high-frequency hearing loss with diagnosed dead regions.

Huss et al. (2001) and Huss and Moore (2005b) obtained pitch matches and octave matches for subjects with high-frequency dead regions. Results for a subject with an extensive high-frequency dead region are shown in Figure 7.3 (results for one ear only; the other ear was “dead”). The dead region was estimated to start at about 1.2 kHz. Pitch matches within one ear (top) were reasonably accurate for frequencies up to 1.25 kHz, and then became much more erratic, indicating that a clear pitch percept was not obtained at frequencies above 1.25 kHz. Octave matches with the lower tone fixed in frequency (middle) resulted in frequency ratios around 2 (the “expected” value) for fixed frequencies up to 0.5 kHz. For a fixed frequency of 1 kHz, the upper tone was adjusted to about 1.4 kHz; when the upper tone fell within the dead region its pitch was higher than “normal.” Octave matches with the upper tone fixed in frequency (bottom) resulted in frequency ratios around (but a little above) 0.5 (the “expected” value) for fixed frequencies up to 1 kHz. For fixed frequencies

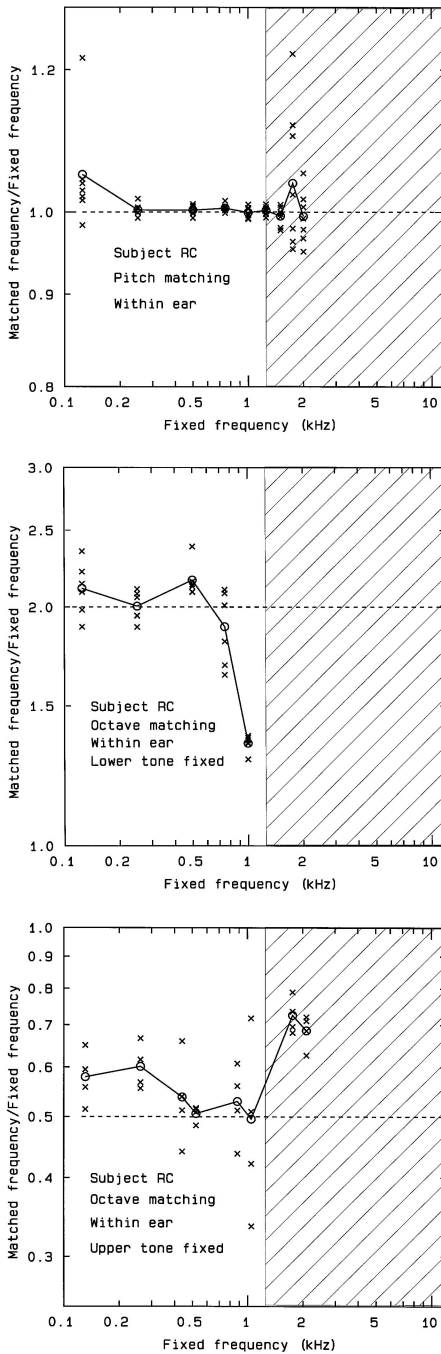


FIGURE 7.3. Results of the pitch-matching and octave-matching tasks within one ear for subject RC, who had a high-frequency dead region starting at about 1.2 kHz (the other ear was “dead”). Matches are shown in the *top panel*. Octave matches were made with the lower tone fixed in frequency (*middle*) and with the upper tone fixed in frequency (*bottom*).

of 1.76 and 2 kHz, octave matches clearly deviated from a ratio of 0.5. For tones whose frequencies fell well within the dead region, the perceived pitch was shifted upwards, although it was also unclear.

Taken together, the results of studies of pitch perception using people with dead regions indicate the following:

1. Pitch matches (of a tone with itself, within one ear) are often erratic, and frequency discrimination is poor, for tones with frequencies falling in a dead region. This indicates that such tones do not evoke a clear pitch sensation.
2. Pitch matches across the ears of subjects with asymmetric hearing loss, and octave matches within ears, indicate that tones falling within a dead region sometimes are perceived with a near-“normal” pitch and sometimes are perceived with a pitch distinctly different from “normal.”
3. The shifted pitches found for some subjects indicate that the pitch of low-frequency tones is not represented solely by a temporal code. Possibly, there needs to be a correspondence between place and temporal information for a “normal” pitch to be perceived (Evans 1978; Loeb et al. 1983; Srulovicz and Goldstein 1983). Alternatively, as noted earlier, temporal information may be “decoded” by a network of coincidence detectors whose operation depends on the phase response at different points along the basilar membrane (Loeb et al. 1983; Shamma and Klein 2000). Alteration of this phase response by cochlear hearing loss (Ruggero et al. 1996) may prevent effective use of temporal information.

## 5. Pitch Anomalies in the Perception of Pure Tones

Although people with low-frequency hearing loss sometimes perceive the pitch of low-frequency tones in a more or less “normal” way, cochlear hearing loss at low or high frequencies does sometimes lead to changes in perceived pitch, even when dead regions are not present. For people with unilateral cochlear hearing loss, or asymmetrical hearing losses, the same tone presented alternately to the two ears may be perceived as having different pitches in the two ears. This effect is given the name *diplacusis*. Sometimes different pitches are perceived even when the hearing loss is the same in the two ears. The magnitude of the shift can be measured by getting the subject to adjust the frequency of the tone in one ear until its pitch matches that of the tone in the other ear.

According to the place theory, cochlear damage might result in pitch shifts for two reasons. The first applies when the amount of hearing loss varies with frequency and especially when the amount of IHC damage varies with characteristic frequency. When the IHCs are damaged, transduction efficiency is reduced, and so a given amount of basilar membrane vibration leads to less neural activity than when the IHCs are intact. When IHC damage varies with characteristic frequency, the peak in the neural excitation pattern evoked by a tone will shift away from a region of greater IHC loss. Hence the perceived pitch is

predicted to shift away from that region. Results from early studies of diplacusis (de Mare 1948; Webster and Schubert 1954) were generally consistent with this prediction, showing that when a sinusoidal tone is presented in a frequency region of hearing loss, the pitch shifts toward a frequency region where there is less hearing loss. For example, in a person with a high-frequency hearing loss, the pitch was reported to be shifted downward. However, there are clearly cases where the pitch does not shift as predicted (see later for examples).

An alternative way in which pitch shifts might occur is by shifts in the position of the peak excitation on the basilar membrane; such shifts can occur even for a flat hearing loss. The tips of tuning curves on the basilar membrane and of neural tuning curves often shift toward lower frequencies when the functioning of the cochlea is impaired by administration of anaesthetic or ototoxic drugs (Sellick et al. 1982; Ruggero and Rich 1991). This means that the maximum excitation at a given place is produced by a lower frequency than normal. Hence, for a given frequency, the peak of the basilar membrane response in an impaired cochlea would be shifted toward the base, that is, toward places normally responding to higher frequencies. This leads to the prediction that the perceived pitch should be shifted upward. Several studies have found that this is usually the case. For example, Gaeth and Norris (1965) and Schoeny and Carhart (1971) reported that pitch shifts were generally upward regardless of the configuration of loss. However, it is also clear that individual differences can be substantial, and subjects with similar patterns of hearing loss (absolute thresholds as a function of frequency) can show quite different pitch shifts.

Burns and Turner (1986) measured changes in pitch as a function of level, by obtaining pitch matches between a tone presented at a fixed level (midway, in decibels, between the absolute threshold and 100 dB SPL) and a tone of variable level. The tones were presented alternately to the same ear. Normally hearing subjects usually show small shifts in pitch with level in this type of task; the shifts are rarely greater than about 3% (Terhardt 1974; Verschuure and van Meeteren 1975). The hearing-impaired subjects of Burns and Turner often showed abnormally large pitch-level effects, with shifts up to 10%. A common pattern was an abnormally large negative pitch shift with increasing level for low-frequency tones.

Burns and Turner (1986) obtained several other measures from their subjects, including PTCs in forward masking, FDLs, measures of diplacusis, and octave judgments. There was a tendency for increased FDLs and increased pitch-matching variability in frequency regions where the PTCs were broader than normal. The exaggerated pitch-level effects occurred both in frequency regions where PTCs were broader than normal, and (sometimes) in regions where both absolute thresholds and PTCs were normal. The results of the diplacusis measurements and octave matches indicated that the large pitch-intensity effects were mainly a consequence of large increases in pitch at low levels; the pitch returned to more "normal" values at higher levels.

As pointed out by Burns and Turner, these results are difficult to explain by the place theory. There is no evidence to suggest that peaks in basilar membrane

responses or in neural excitation patterns of ears with cochlear damage are shifted at low levels but return to “normal” positions at high levels. Also, even in subjects with similar configurations of hearing loss, the pitch shifts and changes in pitch with level can vary markedly. Furthermore, as pointed out earlier, low-frequency pure tones can evoke low pitches even in people who appear to have no IHCs or neurons tuned to low frequencies.

The results are also problematic for the temporal theory. There is no obvious reason why systematic shifts in pitch should occur as a result of cochlear damage or of changes in level. Unfortunately, there seem to be no physiological data concerning the effects of level on phase locking in ears with cochlear damage. As pointed out earlier, it is possible that the central mechanisms involved in the analysis of phase-locking information make use of the propagation time of the traveling wave on the basilar membrane (Loeb et al. 1983; Shamma 1985). This time can be affected by cochlear damage, and this could disrupt the processing of the temporal information by central mechanisms.

In summary, the perceived pitch of pure tones can be affected by cochlear hearing loss, and changes in pitch with level can be markedly greater than normal. Large individual differences occur, even between subjects with similar absolute thresholds. The mechanisms underlying these effects remain unclear.

## 6. Pitch Perception of Complex Tones by People with Cochlear Hearing Loss

### 6.1. *Theoretical Considerations*

As described earlier, cochlear hearing loss is usually associated with reduced frequency selectivity; the auditory filters are broader than normal. This will make it more difficult to resolve the harmonics of a complex tone, especially when the harmonics are of moderate harmonic number. For example, for an  $F_0$  of 200 Hz, the 4th and 5th harmonics would be quite well resolved in a normal auditory system, but would be poorly resolved in an ear where the auditory filters were, say, three times broader than normal.

In the normal auditory system, complex tones with low, resolvable harmonics give rise to clear pitches while tones containing only high, unresolvable harmonics give less clear pitches; see Plack and Oxenham, Chapter 2. Since cochlear hearing loss is associated with poorer resolution of harmonics, one might expect that this would lead to less clear pitches, and poorer discrimination of pitch than normal.

Spectro-temporal theories of pitch perception (see de Cheveigné, Chapter 6) assume that the perception of the pitch of complex tones depends on both place (spectral) analysis and temporal analysis (Meddis and Hewitt 1988; Meddis and O'Mard 1997; Moore 2003). Evidence for a role of the time pattern of the waveform evoked on the basilar membrane by the higher harmonics comes from studies of the effect of changing the relative phase of the components in a



complex tone. Changes in phase can markedly alter both the peak factor of the waveform on the basilar membrane (Ritsma and Engel 1964; Moore 1977; Patterson 1987b) and the number of major waveform peaks per period (Ritsma and Engel 1964; Moore 1977; Patterson 1987a; Moore and Glasberg 1988a; Shackleton and Carlyon 1994). If a complex tone contains low harmonic numbers (say the second, third, and fourth), they will be resolved on the basilar membrane. In this case, the relative phase of the harmonics is of little importance as the envelope on the basilar membrane does not change when the relative phases of the components are altered. However, if the complex tone contains only high harmonics (above about the 8th), then changes in the relative phase of the harmonics can affect both the pitch value and the clarity of pitch (Moore 1977; Patterson 1987a; Houtsma and Smurzynski 1990; Shackleton and Carlyon 1994). It seems likely that pitches based on high unresolved harmonics will be clearest when the waveforms evoked at different points on the basilar membrane each have a single major peak per period of the sound. Given that hearing-impaired subjects have broader-than-normal auditory filters, it can be expected that their perception of pitch and their ability to discriminate repetition rate might be more affected by the relative phases of the components than is the case for normally hearing subjects. For subjects with broad auditory filters, even the lower harmonics would interact at the outputs of the auditory filters, giving a potential for strong phase effects. Changes in phase locking and in cochlear traveling wave phase could also lead to less clear pitches and poorer discrimination of complex tone pitch than normal.

## *6.2. Experimental Studies of Fundamental Frequency Discrimination*

The pitch discrimination of complex tones by hearing-impaired people has been the subject of several studies (Hoekstra and Ritsma 1977; Rosen 1987; Moore and Glasberg 1988b, 1990; Moore and Peters 1992; Arehart 1994; Moore and Moore 2003). Most studies have required subjects to identify which of two successive harmonic complex tones had the higher  $F_0$  (corresponding to a higher pitch). The threshold determined in such a task will be described as a fundamental frequency difference limen ( $F_0DL$ ). As an example, we consider the results of Moore and Peters (1992). They tested four groups of subjects: young subjects with normal hearing; young subjects with impaired hearing; elderly subjects with normal or near-normal hearing; and elderly subjects with impaired hearing. The complex tones were composed of equal-amplitude harmonics with  $F_0$ s of 50, 100, 200, and 400 Hz. Each component had a level of 75 dB SPL, chosen to be above threshold for all subjects. The tones contained harmonics 1 to 12, 6 to 12, 4 to 12 and 1 to 5. The components of the harmonic complexes were added in one of two phase relationships, all cosine phase or alternating cosine and sine phase. The former results in a waveform with prominent peaks and low amplitudes between the peaks. The latter results in a waveform with a much flatter envelope and with two major waveform peaks per period.

The geometric mean values of the F0DLs are plotted separately for each group in Figure 7.4. F0DLs are expressed as a percentage of F0 and plotted on a logarithmic scale. Each symbol represents results for a particular harmonic complex, as indicated by the key in the upper right panel. The results have been averaged across the two phase conditions; phase effects will be discussed later.

Performance was clearly worse for the two hearing-impaired groups than for the young normal-hearing group. F0DLs for the elderly normal-hearing group were also higher than for the young normal-hearing group, especially at low F0s. Indeed, at F0 = 50 Hz, F0DLs for the elderly normal-hearing group were similar to those for the two impaired groups. For all four groups, F0DLs for

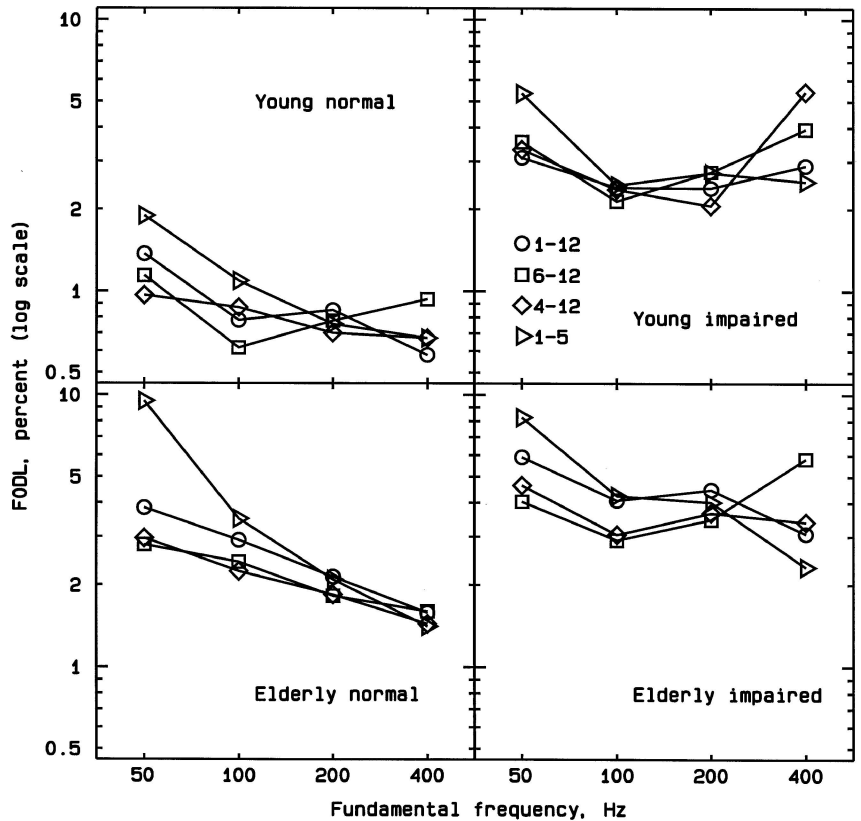


FIGURE 7.4. Mean results of Moore and Peters (1992). The geometric mean values of the DLCs, expressed as a percentage of F0, are plotted separately for each group. Each symbol represents results for a particular harmonic complex, as indicated by the key in the upper right panel. The results have been averaged across two phase conditions, with components added in cosine phase or alternating phase.

$F_0 = 50$  Hz were higher for the complex containing harmonics 1 to 5 than for any of the other complexes. For the two elderly groups and for the lower  $F_0$ s, performance was worse for complex 1 to 12 than for complexes 4 to 12 or 6 to 12, indicating that adding lower harmonics to a complex tone can actually impair performance. This may happen because, when auditory filters are broader than normal, adding lower harmonics can create more complex waveforms at the outputs of the auditory filters, making temporal analysis more difficult (Rosen and Fourcin 1986). Overall, these results suggest that, for low  $F_0$ s, pitch is extracted primarily from harmonics above the 5th. This is consistent with results presented by Moore and Glasberg (1988a, 1990). In contrast, for  $F_0 = 400$  Hz, the complexes containing only high harmonics (6 to 12 and 4 to 12) tended to be the most poorly discriminated, especially by the two impaired groups. These results indicate that the dominant region for pitch is not fixed in harmonic number, but shifts upward in harmonic number as  $F_0$  decreases, as suggested by earlier work (Plomp 1967; Patterson and Wightman 1976; Moore et al. 1985a).

Consider now the effects on the F0DLs of the relative phases of the components. For all four subject groups, F0DLs were, on average, larger for components added in alternating phase than for components added in cosine phase. The phase effect was statistically significant for each subject group. The mean F0DLs for each group, harmonic complex and phase are shown in Figure 7.5;

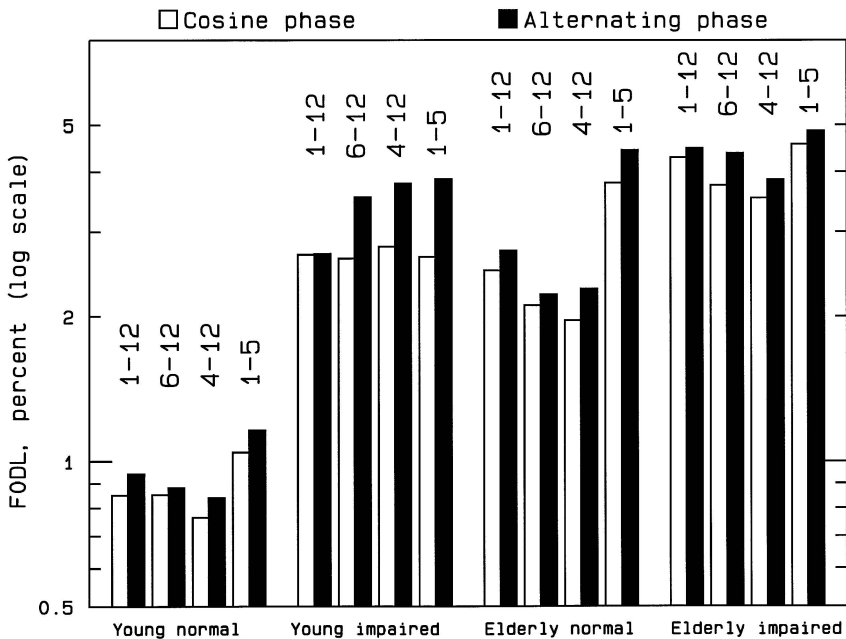


FIGURE 7.5. The mean DLCs for each group tested by Moore and Peters (1992). Results are shown for each harmonic complex and phase, but are averaged across  $F_0$ .

results have been averaged across F0s, since only one group showed a significant interaction of phase with F0. In every case shown, F0DLs are larger for alternating phase than for cosine phase, but the effects overall are rather small. This is somewhat misleading, however, in indicating the influence of phase, since the direction of the effect (whether the change from cosine to alternating phase made performance worse or better) varied in an idiosyncratic way across subjects, F0s, and harmonic contents. Phase effects for individual subjects were often considerably larger than indicated in Figure 7.5.

Overall, studies of F0DLs for subjects with cochlear hearing loss have revealed the following:

1. There was considerable individual variability, both in overall performance and in the effects of harmonic content.
2. For some subjects, when F0 was low, F0DLs for complex tones containing only low harmonics (1 to 5) were markedly higher than for complex tones containing higher harmonics. Since these subjects generally had broader auditory filters than normal, harmonics above the fifth would probably have been unresolved. Hence the pattern of the results suggests that a clearer pitch was conveyed by the unresolved harmonics than by the resolved harmonics.
3. For some subjects, F0DLs were larger for complex tones with lower harmonics (1 to 12) than for tones without lower harmonics (4 to 12 and 6 to 12) for F0s up to 200 Hz. In other words, adding lower harmonics made performance worse. This may happen because, when auditory filters are broader than normal, adding lower harmonics can create more complex waveforms at the outputs of the auditory filters. For example, there may be more than one peak in the envelope of the sound during each period, and this can make temporal analysis more difficult (Rosen and Fourcin 1986; Rosen 1986).
4. The F0DLs were mostly only weakly correlated with measures of frequency selectivity. There was a slight trend for large F0DLs to be associated with poor frequency selectivity, but the relationship was not a close one. Some subjects with very poor frequency selectivity had reasonably small F0DLs.
5. There were sometimes significant effects of component phase. F0DLs tended to be larger for complexes with components added in alternating sine/cosine phase than for complexes with components added in cosine phase. However, the opposite effect was sometimes found. The direction of the phase effect varied in an unpredictable way across subjects and across type of harmonic complex. Phase effects tended to be stronger for hearing-impaired than for normally hearing subjects.
6. Hearing-impaired subjects appear to be less sensitive than normally hearing subjects to the temporal fine structure of complex tones; they appear to reply more on the timing of the envelope than on the timing of the fine structure within the envelope (Moore and Moore 2003).

As noted earlier, it may be the case that the clarity of pitch is greatest, and the F0DL is smallest, when the waveforms evoked at different points on the

basilar membrane each contain a single major peak per period. The basilar-membrane waveforms are determined by the magnitude and phase responses of the auditory filters, and these may vary markedly across subjects and center frequencies depending on the specific pattern of cochlear damage. The variability in the phase effect may arise from variability in the properties of the auditory filters across subjects and across center frequencies.

Overall, these results suggest that, relative to people with normal hearing, people with cochlear damage depend relatively more on temporal information from unresolved harmonics and less on spectral/temporal information from resolved harmonics. The results lend support to spectro-temporal theories of pitch perception. The variability in the results across people, even in cases where the audiometric thresholds are similar, may occur partly because of individual differences in the auditory filters and partly because loss of neural synchrony is greater in some people than others. People in whom neural synchrony is well preserved may have good pitch discrimination despite having broader-than-normal auditory filters. People in whom neural synchrony is adversely affected may have poor pitch discrimination regardless of the degree of broadening of their auditory filters.

### *6.3. Perception of Musical Intervals*

Arehart and Burns (1999) studied the ability of subjects with high-frequency cochlear hearing loss to identify musical intervals between complex tones containing just two harmonics. All subjects were musically trained. The task was similar to that used by Houtsma and Goldstein (1972), and the harmonics were presented at a low sensation level (14 dB SL) either monaurally or dichotically (one harmonic to each ear). To prevent subjects basing their judgments on the pitches of individual harmonics, the rank (harmonic number) of the lowest harmonic in each complex was randomly varied from one stimulus to the next. When the F0 was low and the (mean) harmonic number was low, subjects showed excellent performance. However, for high F0s and high (mean) harmonic numbers, performance worsened markedly and was much poorer than reported by Houtsma and Goldstein (1972) for normally hearing subjects. The highest frequency of the harmonics for which the task was possible was similar for monaural and dichotic presentation. Since resolution of the harmonics should not have been a problem for the dichotic presentation, this finding suggests that some factor other than reduced frequency selectivity limited the ability of the subjects to extract residue pitch from tones with high F0s and high harmonic numbers. Arehart and Burns (1999) suggested that the poor performance of their subjects when the harmonics fell in the region of the hearing loss may have been due to degraded temporal information from that region.

## 7. Perceptual Consequences of Altered Frequency Discrimination and Pitch Perception

### *7.1. Effects on Speech Perception*

Hearing-impaired people generally have a poorer-than-normal ability to understand speech, and altered pitch perception may contribute to this problem. In all languages, the pitch patterns of speech indicate which are the most important words in an utterance, they distinguish a question from a statement and they indicate the structure of sentences in terms of phrases. In “tone” languages, such as Mandarin Chinese, Zulu, and Thai, pitch can affect word meanings. Pitch also conveys nonlinguistic information about the gender, age, and emotional state of the speaker. Supplementing lip reading (speechreading) with an auditory signal containing information only about voice pitch can result in a substantial improvement in the ability to understand speech (Risberg 1974; Rosen et al. 1981; Grant et al. 1985). The use of a signal that conveys information about the presence or absence of voicing (i.e., about whether a periodic complex sound is present or not) gives less improvement than when pitch is signaled in addition (Rosen et al. 1981). It seems likely that reduced ability to discriminate pitch changes, as occurs in people with cochlear hearing loss would reduce the ability to use pitch information in this way.

For complex tones, people with cochlear hearing loss are often more affected by the relative phases of the components than are normally hearing people. When a hearing-impaired person is reasonably close to a person speaking, and when the room has sound absorbing surfaces, the waveforms reaching the listener’s ears when a voiced sound is produced will typically have one major peak per period. These peaky waveforms may evoke a distinct pitch sensation. On the other hand, when the listener is some distance from the speaker, and when the room is reverberant, the phases of the components become essentially random (Plomp and Steeneken 1973) with the result that the waveforms are less peaky. In this case, the evoked pitch may be less clear. The ability of hearing-impaired listeners to extract pitch information in everyday situations may be overestimated by studies using headphones or conducted in rooms with sound-absorbing walls.

For normally hearing listeners, several studies have shown that when two people are talking at once, it is easier to “hear out” the speech of individual talkers when their voices have different F0s (Brokx and Nootboom 1982). This effect may depend on several factors (Culling and Darwin 1993). First, when the F0s of two voices differ, the lower resolved harmonics of the voices (when both are producing voiced sounds, such as vowels) have different frequencies and excite different places on the basilar membrane. This allows the brain to separate the harmonics of the two voices and to attribute to one voice only those components whose frequencies form a harmonic series. This mechanism would be adversely affected by cochlear hearing loss, since reduced frequency selectivity would lead to poorer resolution of the harmonics. Second, when the F0

difference between two voices is small, temporal interactions between the lower harmonics, a form of beats, have the effect that the neural response to the two voices is dominated alternately by first one voice and then the other (Culling and Darwin 1994). The auditory system appears to be able to listen selectively in time to extract a representation of each vowel. This mechanism would also be adversely affected by cochlear hearing loss, since it depends on the interaction of pairs of closely spaced harmonics (one from each voice) that are well separated from other pairs on the basilar membrane. Finally, the higher harmonics would give rise to complex waveforms on the basilar membrane, and these waveforms would differ in repetition rate for the two voices. The brain may be able to use the differences in repetition rate to enhance separation of the two voices (Assmann and Summerfield 1990). This mechanism might depend on the two voices having different short-term spectra. At any one time, the peaks in the spectrum of one voice would usually fall at different frequencies from the peaks in the spectrum of the other voice. Hence, one voice would dominate the basilar membrane vibration patterns at some places, while the other voice would dominate at other places. The local temporal patterns could be used to determine the spectral characteristics of each voice. This mechanism would also be impaired by cochlear hearing loss, for two reasons. First, reduced frequency selectivity would tend to result in more regions on the basilar membrane responding to the harmonics of both voices, rather than being dominated by a single voice. Second, abnormalities in temporal coding might lead to less effective representations of the F0s of the two voices.

The role of F0 differences in enhancing the ability to identify simultaneously presented pairs of vowels has been studied by Arehart et al. (1997). In a double-vowel identification task, normal-hearing listeners showed an 18.5% benefit from an F0 differences of two semitones, while impaired listeners showed a 16.5% benefit. In a second task, subjects were required to identify a target vowel in the presence of a masking vowel; the “threshold” for identification of the target vowel was measured. For normal listeners, the threshold decreased by 9.4 dB with increasing F0 separation, while for impaired listeners the threshold decreased by only 4.4 dB. Overall, the performance of the hearing-impaired listeners was significantly worse than that of the normal listeners. In a later study, Arehart (1998) showed that increasing the audibility of the second and higher formants using high-frequency amplification (25 dB above 1000 Hz) did not improve double-vowel identification by hearing-impaired listeners with F0 differences of zero and two semitones. This suggests that the reduced benefit of F0 differences for the hearing-impaired listeners was not due to an inability to hear the higher formants.

Summers and Leek (1998) measured both thresholds for discrimination of the F0 of (individual) synthetic vowels (F0DLs) and the ability to identify double vowels. Normally hearing listeners and hearing-impaired listeners with small F0DLs obtained benefit when the F0 separation of the two vowels was increased up to four semitones. In contrast, hearing-impaired listeners with large F0DLs did not show any benefit of F0 separation. For a task involving competing

synthetic sentences, the association of F0-based differences in performance and the FODLs was weaker, but, as a group, normally hearing listeners got more benefit from F0 differences than hearing-impaired listeners.

## 7.2. *Effects on Music Perception*

The existence of pitch anomalies (diplacusis and exaggerated pitch-intensity effects) may affect the enjoyment of music. Changes in pitch with intensity would obviously be very disturbing, especially when listening to a live performance, where the range of sound levels can be very large. There have been few, if any, studies of diplacusis for complex sounds, but it is likely to occur to some extent. One person described briefly in Moore (1998), was a Professor of music who had a unilateral cochlear loss. He reported that he typically heard different pitches in his normal and impaired ear, and that musical intervals in his impaired ear sounded distorted. Other subjects have reported that some musical notes do not produce a distinct pitch, and that they get no pleasure from listening to music.

## 8. Introduction to Cochlear Implants

The remainder of this chapter concerns the perception of pitch by users of cochlear implants (see also Zeng et al. 2004). As we shall see, this issue is both clinically important and allows us to address theoretical issues that are difficult to tackle with acoustic hearing. We start with a brief overview of some of the fundamental features of modern cochlear implants.

Cochlear implants represent the world's first successful implanted sensory prosthesis, and have provided hearing to more than 50,000 patients world-wide. The details of the devices differ somewhat between manufacturers, but all modern implants share a number of common features, which are shown schematically in Figure 7.6a:

1. Sound is picked up by a microphone worn behind the ear, and the analog waveform passed to an external speech processor.
2. The processor determines the pattern of stimulation to be applied to the intracochlear electrodes.
3. The output of the speech processor modulates a radio-frequency carrier, which is transmitted across the skin using a coil, and decoded by a receiver-stimulator implanted under the skin.
4. The receiver-stimulator sends electrical signals to each of several electrodes implanted in the cochlea. The electrodes are arranged along the length of the basilar membrane, so that the higher frequency bands of the input waveform control stimulation of the more basal electrodes, with progressively lower bands determining the stimulation on more apical electrodes.

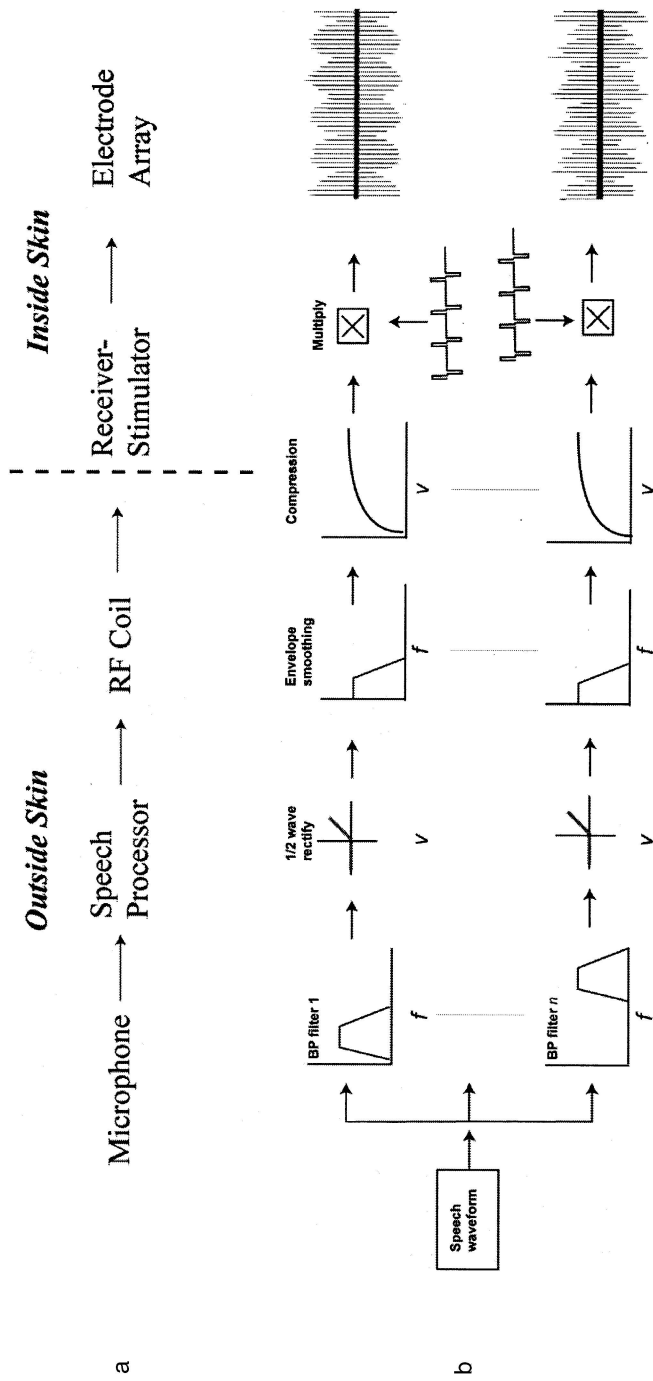


Several different speech-processing algorithms have been implemented. The Continuous Interleaved Sampling (“CIS”) algorithm (Wilson et al. 1991; see our Figure 7.6b) is a good example, because it is fairly straightforward, shares many features with other popular algorithms, and has been implemented by all of the major cochlear implant manufacturers. After high frequencies have been boosted by an initial pre-emphasis, the input is passed through a bank of (typically six to eight) bandpass filters, and the envelope of each filter output is obtained via rectification and low-pass filtering. The envelope from each band is compressed and then modulates a high-rate train of biphasic pulses on the appropriate electrode. The pulses are presented continuously, but are interleaved between the different electrodes, thereby giving the algorithm its name. Figure 7.6c shows the output of the CIS algorithm in response to a synthetic vowel having an F0 of 100 Hz, formant frequencies of 400 and 1200 Hz, and a carrier pulse rate of 813 pps. It can be seen that the F0 of the vowel is conveyed by regular 100-Hz fluctuations in the output of some channels; this is indicated by arrows for the channel centered on 2320 Hz. Geurts and Wouters (2001) have shown that implant patients can use this cue to detect F0 differences between synthetic vowels processed by the CIS algorithm.

## 9. Limitations in the Peripheral Encoding of Sound by Cochlear Implants

Although cochlear implants can be very successful in restoring speech perception, at least in quiet (Skinner et al. 1994), they clearly do not reproduce the pattern of auditory nerve activation that an acoustic stimulus elicits in a normally hearing ear. The differences arise from a combination of surgical limitations, the nature of the hardware, the electrical conductivity of cochlear fluids and structures, and from the speech-processing algorithms used. They are summarized in the following list; where appropriate, we compare them to the effects of sensory hearing loss on peripheral coding (Section 2).

1. The electrode array is usually not fully inserted into the cochlea. Consequently, to convey the full range of frequencies available to a normal listener, each frequency band is usually encoded on an electrode innervating a region of the BM basal to that which would be excited “naturally.” For example, Ketten et al. (1998), in a study of 20 patients implanted with a Cochlear Corporation device, found that the most apical electrode occupied a location with a best frequency, according to Greenwood’s (1990) equation, ranging from 387 Hz to 2596 Hz. Typically, this electrode will convey information about frequencies below 240 Hz, leading to a frequency-to-place mismatch that can be greater than three octaves.
2. There will be degeneration of the auditory nerve innervating all regions of the cochlea, with the dendrites being more vulnerable than the cell bodies



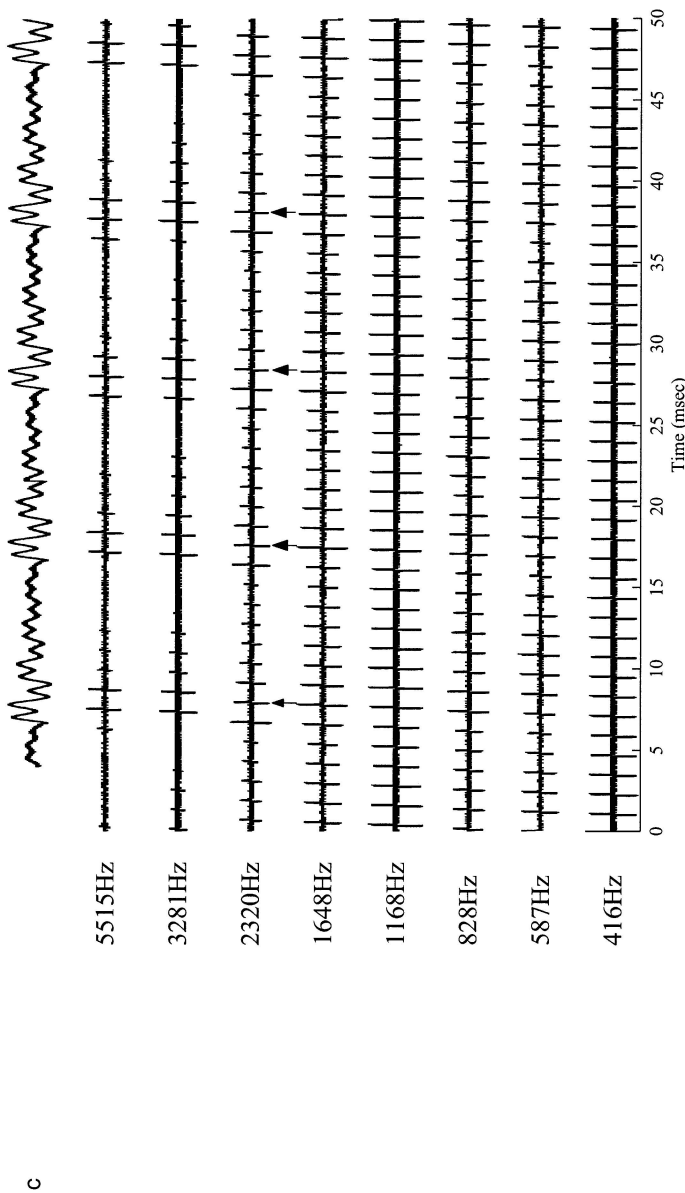


FIGURE 7.6. Part (a) shows a schematic representation of the various components of a modern cochlear implant. In many devices, the microphone and speech processor are housed in a common unit, worn behind the ear, resembling a behind-the-ear hearing aid. Part (b) is a schematic illustration of an  $n$ -channel CIS algorithm, in which the initial high-frequency preemphasis has been omitted for clarity. Only the lowest and highest channels are shown. The time scale of the output waveforms is much coarser than that of the biphasic “carrier” trains illustrated. Part (c) shows the output of an eight-channel CIS algorithm, with the center frequency of each channel shown on the left. The stimulus was a synthetic /a/ having an F0 of 100 Hz, with harmonics summed in sine phase. The 100-Hz fluctuations in the output of the channel centered on 2320 Hz are indicated by arrows; similar fluctuations can be seen in some other channels. Only two formants were synthesized, having frequencies of 400 and 1200 Hz, respectively.

and axons (for a review, see Shepherd and Javel 1997). This may result in “dead” regions of the cochlea, where the degeneration is complete. As with the dead regions described earlier for impaired acoustic hearing (Section 2), stimulation applied to one part of the cochlea may consequently be conveyed to the brain by auditory nerve (AN) fibers innervating another part.

3. In acoustic hearing, the AN phase locks to resolved frequency components. Hence the temporal pattern of firing and the place of excitation are to some extent linked. For example, if the frequency of a 500-Hz pure tone is increased by 10%, there is a corresponding shift in both the pattern of phase locking and in the subset of AN fibers that respond to the tone. However, the most widely used cochlear implant speech-processing strategies apply pulse trains having the same rate to each electrode channel. It may be that a correspondence between the temporal and place-of-excitation cues to frequency is important for pitch perception (Loeb et al. 1983; Carlyon and Deeks 2002). This idea receives some support from the evidence reviewed in Section 4, showing that, in impaired acoustic hearing, tones falling in a dead region often have a weak pitch, perhaps due to a mismatch between place and phase-locking cues.
4. In normal hearing, there is a phase transition around peaks of the traveling wave (Kim et al. 1980; Dallos et al. 1996), and this may be important for pitch perception based on timing comparisons between different parts of the auditory nerve fiber array (Loeb et al. 1983; Shamma and Klein 2000). These phase transitions are not encoded by cochlear implant speech processors. As pointed out in Section 2, the transitions may also be disrupted in acoustic hearing by sensory hearing loss.
5. Frequency selectivity *may* be reduced. Chatterjee and Shannon (1998) measured forward masked excitation patterns in four users of the Nucleus Corporation 22-channel implant. They compared the resulting patterns to analogous measurements obtained with acoustic stimulation of a normally hearing listener, with acoustic frequency converted to electrode position using Greenwood’s (1990) formula. For two of the implanted subjects, the excitation patterns were slightly broader than normal, whereas one showed a spatial extent that was more than twice as wide. A fourth showed excitation patterns that were sharp near the tip but which, for some electrode pairs, were nonmonotonic at wider masker-probe separations.

## 10. Place and Timing Cues Conveyed by Electrical and Acoustical Stimulation

As described in Sections 3.1 and 9, the peripheral encoding of a pure tone or resolved partial in acoustic hearing may involve both timing (“phase locking”) and place-of-excitation cues. This can make it hard to tell which cue is being used in a given experimental task, and is at least partly responsible for the decades of debate over the importance of place and timing cues in pitch per-

ception. In contrast, by replacing the microphone and speech processor in Figure 7.6a with an experimental device controlled by a computer, cochlear implant researchers can control these two parameters independently. For example, by varying the rate or temporal pattern of current pulses applied to a single electrode, one can study temporal pitch perception while holding place-of-excitation constant. Conversely, temporal cues can be held constant by applying the same temporal pattern of pulses to different electrodes, thereby allowing one to study the effects of changing only the place of excitation.

There is some evidence that, although the place-of-excitation and timing manipulations can both be described as affecting pitch, they tap quite separate perceptual dimensions. Tong et al. (1983) presented subjects with a total of nine different stimuli, each of which consisted of one of three different pulse rates applied to one of three electrodes. On each trial, three pairs of stimuli were presented, and subjects were required to report which pair sounded the most dissimilar. The resulting difference scores were then analyzed using multidimensional scaling, which resulted in a two-dimensional solution, where the dimensions corresponded to rate and place. In a more recent study (McKay et al. 2000), subjects were required to detect a shift in place of excitation, a change in pulse rate, or a combination of the two. Subjects reliably detected the combined changes better than changes in either the rate or place-of-excitation alone. McKay et al. reasoned that if place and rate changes were combined into a single perceptual dimension, then performance would be better when the two changes to be combined were “consistent,” in that they produced a pitch shift in the same direction (e.g., rate increase + basal shift, or rate decrease + apical shift), than when they were “inconsistent” (e.g., rate increase + apical shift, or decrease + basal shift). No such difference was found.

The evidence described above suggests that place and timing cues to pitch are indeed independent for the stimuli used in most cochlear implant experiments. However, we should note that this does not rule out the possibility that a match between place of excitation and phase locking is important for pitch. In the experiments we have described, the mismatch between rate and place of stimulation is likely to have been very large, and it is probable that no conditions were included in which a close match was obtained. What those experiments do demonstrate, however, is that rate and place cues are independent for the stimuli studied (and probably for most others). Consequently, implant experiments provide a golden opportunity for studying timing and place-of-excitation cues in isolation.

### *10.1. Temporal Codes*

When the rate of a train of electrical pulses is raised, subjects reliably report hearing an increase in pitch, even when the place of stimulation is held constant. This is generally true only for rates above about 50 pps, below which subjects report a “rattling” percept. When the baseline rate is above 50 pps, but no more than about 200 pps, discrimination is quite good, but variable across subjects.

For example, a sample of 19 subjects taken from five studies (Pfungst et al. 1994; van Hoesel and Clark 1997; McKay et al. 1999, 2000; Zeng 2002) showed that implant users could, on average, detect a 7.3% increase in the rate of a 100-pps pulse train. As with most measurements obtained with implant users, there was a large range of overall performance across subjects, with the lowest threshold being less than 2% and the highest about 18%. Some of this variability probably stems from differences in procedure across studies; for example, McKay et al., who roved level from presentation to presentation, obtained higher overall thresholds than Pfingst et al., who did not rove level. However, thresholds also vary substantially between subjects within a single study. An analysis of the data of Pfingst et al. shows that the rate DLs obtained at a comfortable listening level correlated significantly with length of deafness prior to implantation ( $r = 0.97$ ,  $df = 4$ ,  $p < 0.01$ ). For this fairly small sample of five subjects, then, duration of deafness can account for a substantial portion (74%) of the variance.

At higher overall rates, performance deteriorates dramatically, and most patients are unable to detect a rate increase for baseline rates above about 300 pps (Shannon 1983; Tong and Clark 1985; Townshend et al. 1987; McKay et al. 2000; Zeng 2002). Again, there is substantial inter-listener variation, and there are reports of a few implant users being able to detect rate increases for rates as high as 1000 pps (Townshend et al. 1987; Wilson et al. 1997). Similar findings have been obtained for sinusoidal electrical stimulation (Fourcin et al. 1979; Shannon 1983).

There is reasonably strong evidence that temporal cues, on their own, can elicit a sense of musical pitch (Fourcin et al. 1979). Pijl and Schwarz (1995) required three implant users to identify simple melodies, picked from a closed set of eight, and played on a single channel of their implant. Pitch was encoded solely by changes in pulse rate. The duration of each note and the silent gaps between notes were held constant in order to eliminate possible rhythm cues. Despite this, when the lowest note in each melody was played at 75 pps, performance was at ceiling for all three subjects. Performance deteriorated at higher pulse rates, consistent with the deterioration observed in the discrimination data described earlier. Even more convincingly, subjects could identify whether the musical interval between two notes was sharp, flat, or in tune relative to a specified interval (e.g., “a minor 3rd,” “a 5th”). A similar ability was exhibited by an implant user tested by McDermott and McKay (1997), who, prior to his deafness, had been trained as a piano tuner. He could adjust the pulse rate applied to one electrode, so that the resulting pitch formed a pre-specified musical interval relative to a preceding stimulus applied to that same electrode.

Although implant users are able to extract musical pitch from a purely temporal code, there is evidence that they cannot do so as effectively as do normally hearing listeners. One hint comes from the discrimination thresholds of about 7% at low baseline rates (Pfungst et al. 1994; van Hoesel and Clark 1997; McKay et al. 1999, 2000; Zeng 2002), which are considerably higher than the FDLs for acoustic pure tones. As described by Plack and Oxenham (Chapter 2), there is

good evidence that normally hearing listeners encode the frequencies of pure tones with low frequencies using phase-locking cues, and FDLs for tones around 100 Hz are typically below 1% (Moore 1973b). Perhaps more strikingly, the evidence that normal listeners use phase locking up to about 4 to 5 kHz (Moore 1973b; Micheyl et al. 1998; Moore and Sek 1996) contrasts sharply with the marked deterioration in rate discrimination for electric pulse trains above about 300 pps. This latter paradox was recently investigated by Carlyon and Deeks (2002), who considered two hypotheses. One was that the high-rate deterioration observed in electric hearing could be due to the mismatch between place and rate of stimulation (with pulse rates of a few hundred pulses per second being applied to regions of the cochlea normally tuned to frequencies of several thousand Hertz). The other was that, although phase-locking to electric pulse trains can be observed up to very high rates in recently deafened animals, this is not the case for most human implant users, who may have been deaf for several years and who are likely to be stimulated at a lower current level than in animal experiments (van den Honert and Stypulkowski 1987; Shepherd and Javel 1997).

Carlyon and Deeks (2002) used a simulation of electric hearing in which an acoustic pulse train, having a rate of a few hundred pps, was passed through a fixed bandpass filter centered on a frequency of several thousand Hertz. To avoid resolvable harmonics, Carlyon and Deeks used filtered harmonic complexes whose components were, in one condition, summed in alternating phase. This stimulus, which resembles a pulse train, allows one to double the pulse rate relative to that for a sine-phase complex without altering the spacing of the components. Carlyon and Deeks reasoned that, if the limitation observed with implant users was entirely due to the mismatch between place and rate of stimulation, then a similar limitation should be observed with this acoustic simulation. Contrary to this prediction, when the pulse trains were filtered between 7800 and 10,800 Hz, all the normally hearing listeners could perform rate discrimination at a pulse rate as high as 600 pps. Furthermore, at lower pulse rates, DLs were lower than those typically observed with implant users.

Carlyon and Deeks concluded that, for most implant users, the limitation on rate discrimination did not result entirely from a central pitch mechanism being unable to process temporal information effectively when the place and rate of stimulation are mismatched. Because this mismatch did not abolish rate discrimination for normal listeners until the baseline rate reached 600 pps, they argued that rate discrimination by implant users, which is usually impossible at baseline rates above 300 pps, is mediated by a peripheral deficit. However, they also found evidence that, for normal listeners, there *is* a central factor that places an upper limit on rate discrimination at high overall rates. One source of this evidence came from an experiment which was performed using a pulse rate sufficiently high for performance to be at chance with monaural presentation, but which allowed subjects to use a binaural cue. This was achieved by requiring subjects to discriminate between two successive stimuli differing in rate and presented to the left ear, while a copy of the lower-rate stimulus was pre-

sented simultaneously to the right ear. When the left ear also received the lower-rate stimulus, subjects heard a single sound in the middle of the head. However, when the higher-rate signal was presented to the left ear, subjects heard a more diffuse binaural image, which was very easily discriminated from the single, centered image. Hence, even though the right-ear stimulus provided no new information, being the same on all presentations, it resulted in a dramatic improvement in performance. Carlyon and Deeks concluded that information must have been available in the auditory nerve that was accessible to a binaural mechanism, but inaccessible to the temporal pitch mechanism.

## 10.2 *Place of Excitation*

If one applies the same temporal pattern of stimulation to progressively more basal electrodes, implant users reliably report an increase in pitch. For example, Nelson et al. (1995) presented 14 users of the Nucleus CI 22 device with 500-ms pulse trains, applied sequentially to each of two electrode channels. The listeners' task was to state which presentation produced the higher pitch, and an answer was scored as correct if this corresponded to the more basal channel. The results for each pair were converted to the sensitivity index  $d'$ , and, by measuring discrimination for a large number of different electrode pairs, Nelson et al. were able to measure a "cumulative  $d'$ " score for each electrode, relative to the most apical one in the array. The results showed a monotonic increase in pitch with more basal stimulation for all subjects, with very few instances of pitch reversals along the array. However, the variation in sensitivity across subjects was very large; one measure of the average change in sensitivity for each millimeter of electrode separation ranged from 0.12  $d'/\text{mm}$  to 3.16  $d'/\text{mm}$ . The adaptive procedures used to measure thresholds often converge on a value of  $d' = 0.78$ , so the most sensitive listener in this study would have had a threshold of about 0.25 mm. Greenwood's (1990) equation reveals that this is roughly equivalent to the change in place produced by a 3% change in the frequency of an acoustic pure tone. In comparison, FDLs for trained normally hearing listeners vary with frequency, but can be as low as 0.2% at 1000 Hz (Moore 1973b). The threshold value for a median cochlear implantee was about 1.2 mm, corresponding to an approximately 21% change in acoustic frequency.

A demonstration that pitch can be conveyed purely by place-of-excitation cues would disprove models which propose that timing cues are essential for pitch perception (Schouten 1940; Meddis and Hewitt 1991; Patterson et al. 1995). However, it is important to demonstrate that the cue that subjects were using in these experiments really corresponded to pitch. In the study of Nelson et al., and those of others (Townshend et al. 1987; McDermott and McKay 1994; Zwolan et al. 1997), care was taken to ensure that subjects were not basing their responses on the differences in loudness that can result from changes in electrode position. In particular, the absence of feedback in some studies (McDermott and McKay 1994; Nelson et al. 1995) suggests that the basis for performance was some perceptual dimension that subjects can spontaneously and consistently



label as pitch. However, it is possible that this dimension can only be *loosely* defined as pitch (e.g., “that subjective attribute of sound which admits a rank ordering from low to high”—Ritsma 1963). For example, in acoustic hearing, spectrally shaping a noise so that the high-frequency components are relatively more intense may be sufficient for musically inexperienced subjects to report an “increase in pitch” (relative to white noise), but few would argue that this sort of manipulation could convey a convincing melody. A similar sort of thing may be happening when one varies the place of stimulation in a cochlear implant.

To our knowledge, there has only been one attempt to determine whether, in cochlear implant users, place-of-excitation can, by itself, convey musical pitch. McDermott and McKay (1997) presented a musically trained implant user with pulse trains applied sequentially to two different electrodes, and asked him to identify the musical interval between the two sound sensations. He reliably identified larger electrode separations as corresponding to larger musical intervals (Fig. 7.7), and this also occurred, although to a lesser degree, in an extra condition where the more basal electrode was stimulated with a lower pulse rate. However, the function relating the reported interval to electrode separation was significantly shallower than that predicted from the position of the electrodes within the cochlea, on the basis of Greenwood’s (1990) frequency-to-place map. This does not *prove* that place cues cannot convey an accurate sense of musical pitch, as it is possible that regions of AN loss may have caused individual electrodes to excite AN fibers innervating rather different locations on the basilar membrane. However, it does prevent us from concluding that place cues alone *can* support musical interval recognition.

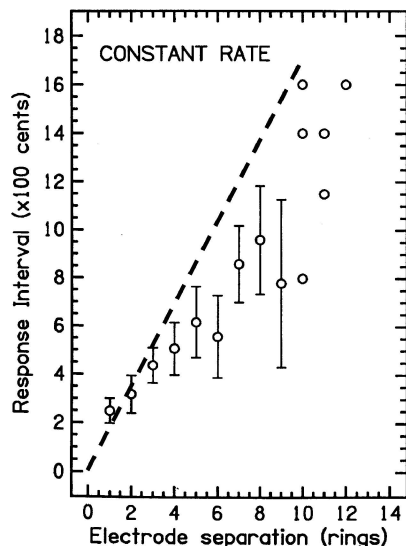


FIGURE 7.7. The musical intervals between two successive stimuli reported by the implant user in McDermott and McKay’s (1997) study, plotted as a function of electrode separation, with associated standard errors. The stimuli were 200-pps electrical pulse trains presented sequentially to two electrodes separated by the number of “rings” shown on the abscissa. (In the device used in that study, the electrode rings are 0.75 mm apart.) The *dashed diagonal line* shows the responses that would be predicted from the positions of the electrodes within the cochlea (see text for details).

In conclusion, although some implant users can detect fairly small changes in place of excitation, and these changes can be described along a dimension of “low to high” or “dull to sharp,” it is not known whether the percepts conveyed meet a strict definition of musical pitch. The fact that purely temporal cues can convey pitch, combined with the evidence that the percepts elicited by place-of-excitation and timing cues are to some extent independent (Tong et al. 1983), suggests that this may not be the case.

## 11. Perceptual Consequences of Altered Pitch Perception in Electric Hearing

### *11.1 Effects on Speech Perception in Quiet*

As discussed in Section 7.1, pitch perception is important in western languages for providing prosodic information and for conveying non-linguistic information concerning the age, gender, and emotional state of the speaker. Perhaps more importantly, pitch provides a useful supplement to speech-reading, and, in “tone” languages, can affect the meanings of words.

As discussed by Plack and Oxenham (Chapter 2), normally hearing listeners derive pitch information primarily from the lower (resolved) harmonics, whose frequencies are each encoded by a separate subset of AN fibers. It is widely believed that phase locking to the frequency of each harmonic contributes to this process. Unfortunately for implant users, the speech processing algorithms used in modern implants are unlikely to result in this information being encoded in the AN. In strategies such as CIS and “SPEAK” (McDermott et al. 1992), a pulse train of the same rate is applied to all electrodes. In addition, and for all strategies, the input filters used to create “channels” are typically too broad to resolve individual harmonics. Furthermore, spread of current along and across the cochlea would increase the “mixing” of harmonics within individual AN fibers, even if devices were modified to have a larger number of electrodes each encoding a narrow range of frequencies. It therefore seems that implant users cannot extract the pitch of an electric input in a way analogous to that used by normally hearing listeners for resolved harmonics. This is likely to impose restrictions on their ability to hear the pitch of a single voice, even when presented in quiet.

As described in Section 10.1, one way in which implant users *can* derive pitch is from the temporal pattern of stimulation. The experiments we described in that section mostly varied the rate of an equal-amplitude pulse train, but one can also manipulate pitch by amplitude modulating a high rate carrier by a low-frequency (e.g., 100 pps) waveform (Fourcin et al. 1979; McKay et al. 1994, 1995; McDermott and McKay 1997; Geurts and Wouters 2001). This is important because, as shown in Figure 7.6c, the pitch of a periodic sound such as a vowel is represented in the CIS algorithm by the frequency at which a high-rate pulse train is modulated. The modulation occurs at a rate equal to  $F_0$ ,

and arises from the beating between many harmonics within each analysis filter. In many ways this is similar to the pitch conveyed by unresolved harmonics in acoustic hearing, and to the type of temporal cue relied on by listeners having cochlear hearing loss (Section 7). Here, also, the scrambling of component phases by room acoustics is likely to make matters worse. The outputs shown in Figure 7.6c were obtained with all harmonics of a vowel synthesized in sine phase; further simulations with random-phase components showed that this reduced the modulation depth in some channels of the algorithm output. Perhaps unsurprisingly, then, patients' perception of the pitch of sounds passed through cochlear implant speech processors is often disappointing. For example Ciocca et al. (2002) found that early-deafened Cantonese-speaking implant users had great difficulty in extracting the pitch information needed to accurately identify Cantonese lexical tones.

### *11.2 Effects on Speech Perception in the Presence of Competing Sounds*

Normally hearing listeners can use differences in F0 to perceptually separate the voices of competing speakers (e.g., Brokx and Nootboom 1982; see also Section 7.1). As discussed by Darwin (Chapter 8), their ability to separate two formants originating from different talkers is greatest when the harmonics of both formants are resolved by the peripheral auditory system (Darwin 1992). We have argued that there is unlikely to be an analogue of this form of processing in electric hearing, and so implant users must rely on other forms of coding to segregate concurrent sounds.

One particularly interesting situation arises when two formants of different voices have similar center frequencies. This will result in a complex pattern of modulation on one or more electrode channels, and it has been suggested that the auditory system could, within each channel, extract the two underlying periodicities—for example, via a “neural cancellation filter” (Assmann and Summerfield 1990; de Cheveigné 1993). Carlyon and co-workers have investigated the perception of pitch when two periodicities are represented in the same channel (Carlyon 1996; Carlyon et al. 2002; Long et al. 2002). Two equal-amplitude pulse trains of different rates were mixed and either applied to a single electrode of a cochlear implant, or passed through a fixed bandpass filter and played acoustically to normally hearing listeners. An example of two individual pulse trains and of the mixture is illustrated schematically in Figure 7.8. The results showed that neither group of listener could analyze the mixture and extract the two underlying periodicities. Instead, they heard a single pitch corresponding to that of the higher-rate pulse train. Carlyon et al. (2002) showed that this and a range of other findings (Carlyon 1997; Plack and White 2000) could be explained by a model in which pitch was derived from a weighted sum of the first-order intervals (those between adjacent pulses) in the stimulus (cf. Chapter 6). In the case of the mixture shown in Figure 7.8, there were no first-order intervals corresponding to the lower-rate stimulus (dashed lines), because these

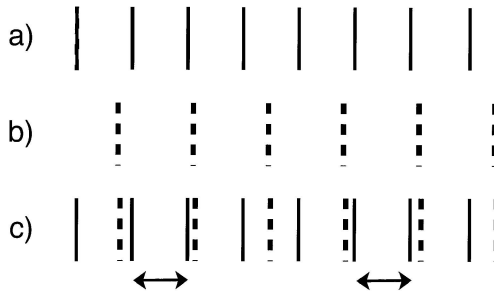


FIGURE 7.8. Parts (a) and (b) show two isochronous pulse trains of different rates. Part (c) shows a schematic of the mixture used by Carlyon and colleagues (Carlyon et al. 2002). The first-order intervals between the pulses from the higher-rate train (*solid lines*) are indicated by *arrows*.

intervals were so long that there was always an intervening pulse from the higher-rate pulse train. In contrast, the most common first-order interval corresponded to that between the pulses from the high-rate stimulus (*solid lines*); two such intervals are indicated by arrows in the figure.

The results obtained by Carlyon and his colleagues suggest that, although listeners may derive a pitch from purely temporal cues, such cues by themselves are unlikely to be sufficient to help in concurrent sound segregation. However, there is some evidence from acoustic simulations that fairly large ( $>10\%$ ) differences in F0 *can* provide a basis for sound segregation, even when encoded only by temporal cues, provided that the two periodicities are represented in *different* populations of AN fibers (Darwin 1992; Carlyon 1994). This situation would arise, for example, when the formants of two competing speakers occupy distinct and well-separated frequency regions. The ability of implant users to exploit F0 differences is therefore likely to depend on the extent to which the spectrum of competing voices, and hence the electrode channels stimulated by them, differs. Another limitation of the F0 cue when more than one source is present is suggested by Figure 7.6c, which reveals that the outputs of channels (center frequencies 416 and 1168 Hz) close to the formant frequencies are not very deeply modulated. Two factors contribute to this: (1) many devices, such as the Advanced Bionics implant used to generate Figure 7.6c, apply a compressive nonlinearity which reduces the modulation depth in channels where there is the most sound energy, and (2) the outputs of channels centered on a formant can be dominated by one or two high-amplitude harmonics, whereas a large modulation depth requires the interaction of many harmonics of approximately equal amplitude. Hence, listeners may depend on the outputs of channels with lower overall levels of stimulation to extract F0, and these will be susceptible to masking by competing sources. The idea that implant users may not be able to exploit F0 differences in segregating competing voices played through their speech processors is supported by the recent finding that, unlike normal

listeners, they do not benefit from a difference in gender between target and interfering speech (Nelson and Jin 2002).

## 12. Summary

In this chapter we have described pitch perception by listeners having cochlear hearing loss and by users of cochlear implants. The study of both clinical populations not only informs the design of effective prostheses, but also allows one to address important theoretical issues. For example, the existence of dead regions in listeners with cochlear hearing loss provides a crucial dissociation of place-of-excitation and phase-locking cues to pitch. A similar dissociation occurs in cochlear implants, where one can independently vary the place and rate of stimulation. Both of these paradigms indicate that a time code is crucially important for pitch perception, but they are also consistent with the proposal that the perception of clear musical pitches requires a close match between place and rate of stimulation (Evans 1978; Loeb et al. 1983; Shamma 1985; Shamma and Klein 2000).

Another similarity between the two patient groups is connected with the harmonics that determine the perceived pitch of complex tones. In both cases, there is a shift away from the pattern observed with normal listeners, for whom resolved harmonics dominate the percept. Instead, there is an increased reliance on neural channels that respond to a mixture of (unresolved) harmonics; the harmonics produce beats at a rate equal to  $F_0$ . For listeners with cochlear hearing loss, this may partly arise from a broadening of the auditory filters, which can mean that even low harmonics are poorly resolved. It may also arise from poor encoding of the lower harmonics, due to mismatch between place and temporal information. For cochlear implantees, the reliance on beating harmonics is due to the relatively broad analysis filters used in speech processors, combined with the spread of electrical charge along and across the cochlea. Although the reasons are different, the consequences are likely to be very similar: elevated discrimination thresholds, increased sensitivity to differences in phase between harmonics, and a reduced ability to use  $F_0$  differences to separate competing sounds. Furthermore, recent evidence that the extraction of pitch from unresolved harmonics is “sluggish” suggests that both groups of listeners may have difficulty in tracking rapid changes in the pitch of a sound (Plack and Carlyon 1995; Micheyl et al. 1998; Carlyon et al. 2000). Finally, it is interesting to speculate on how likely it is that pitch perception can be improved in these two clinical groups, and on the most probable means of achieving that goal. One approach would be to attempt to deliver the auditory signal in a way that allows the impaired auditory system to resolve individual harmonics. For patients with a cochlear hearing loss, this may be a tall order. Although attempts to improve spectral resolution by “sharpening” the spectrum have met with some success (Simpson et al. 1990; Baer et al. 1993), this is likely to reflect the improved resolution of formants, rather than of individual harmonics. For im-

plant users, improvements in the design and placement of electrode arrays have at least the potential for improving frequency resolution. However, it seems unlikely that the effective number of separate electrodes will become sufficient to convey appropriate temporal (and possibly place) information about individual, low-numbered harmonics, regardless of their frequency. An alternative that is perhaps more promising is to improve the pitch percept conveyed by unresolved harmonics. For cochlear implants, there is some evidence that the addition of low levels of noise can improve temporal coding (Zeng et al. 2000; Chatterjee and Robert 2001). In addition, it may be worthwhile to explore signal-processing algorithms (e.g., Geurts and Wouters 2001) that enhance modulations at a rate equal to F0, and to determine whether such algorithms can enhance pitch perception of natural speech sounds and in noisy environments. Although we may not be able to reproduce the neural representation of resolved harmonics that occurs in healthy ears, the enhancement of such modulations may optimize the one form of F0 encoding from which hearing-impaired listeners and implant users *can* benefit.

*Acknowledgments.* We thank Cathy Arehart, Colette McKay, and Christopher Long for helpful comments on a previous version of this chapter. Christopher Long also produced the CIS outputs plotted in Figure 7.6c, which were obtained using a test implant loaned by Patrick Boyle of Advanced Bionics.

## References

- Arehart KH (1994) Effects of harmonic content on complex-tone fundamental-frequency discrimination in hearing-impaired listeners. *J Acoust Soc Am* 95:3574–3585.
- Arehart KH (1998) Effects of high-frequency amplification on double-vowel identification in listeners with hearing loss. *J Acoust Soc Am* 104:1733–1736.
- Arehart KH, Burns EM (1999) A comparison of monotic and dichotic complex-tone pitch perception in listeners with hearing loss. *J Acoust Soc Am* 106:993–997.
- Arehart KH, King CA, McLean-Mudgett KS (1997) Role of fundamental frequency differences in the perceptual separation of competing vowel sounds by listeners with normal hearing and listeners with hearing loss. *J Speech Lang Hear Res* 40:1434–1444.
- Assmann PF, Summerfield AQ (1990) Modeling the perception of concurrent vowels: vowels with different fundamental frequencies. *J Acoust Soc Am* 88:680–697.
- Baer T, Moore BCJ, Gatehouse S (1993) Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: effects on intelligibility, quality and response times. *J Rehab Res Dev* 30:49–72.
- Brokx JPL, Noolteboom SG (1982) Intonation and the perceptual separation of simultaneous voices. *J Phonet* 10:23–36.
- Burns EM, Turner C (1986) Pure-tone pitch anomalies. II. Pitch-intensity effects and diplacusis in impaired ears. *J Acoust Soc Am* 79:1530–1540.
- Carlyon RP (1994) Detecting pitch-pulse asynchronies and differences in fundamental frequency. *J Acoust Soc Am* 95:968–979.

- Carlyon RP (1996) Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker. *J Acoust Soc Am* 99:517–524.
- Carlyon RP (1997) The effects of two temporal cues on pitch judgements. *J Acoust Soc Am* 102:1097–1105.
- Carlyon RP, Deeks JM (2002) Limitations on rate discrimination. *J Acoust Soc Am* 112:1009–1025.
- Carlyon RP, Moore BCJ, Micheyl C (2000) The effect of modulation rate on the detection of frequency modulation and mistuning of complex tones. *J Acoust Soc Am* 108:304–315.
- Carlyon RP, van Wieringen A, Long CJ, Deeks JM, Wouters J (2002) Temporal pitch mechanisms in acoustic and electric hearing. *J Acoust Soc Am* 112:621–633.
- Chatterjee M, Robert ME (2001) Noise enhances modulation sensitivity in cochlear implant listeners: stochastic resonance in a prosthetic sensory system? *J Assoc Res Otolaryngol* 2:159–171.
- Chatterjee M, Shannon RV (1998) Forward masked excitation patterns in multielectrode electrical stimulation. *J Acoust Soc Am* 103:2565–2572.
- Ciocca V, Francis AL, Aisha R, Wong L (2002) The perception of Cantonese lexical tones by early-deafened cochlear implantees. *J Acoust Soc Am* 111:2250–2256.
- Culling JF, Darwin CJ (1993) Perceptual separation of simultaneous vowels: within and across-formant grouping by F0. *J Acoust Soc Am* 93:3454–3467.
- Culling JF, Darwin CJ (1994) Perceptual and computational separation of simultaneous vowels: Cues arising from low-frequency beating. *J Acoust Soc Am* 95:1559–1569.
- Dallos P, Popper R, Fay R (1996) *The Cochlea*. New York: Springer-Verlag.
- Darwin CJ (1992) Listening to two things at once. In: Schouten MEH (ed), *The Auditory Processing of Speech—From Sounds to Words*. Berlin: Mouton de Gruyter, pp. 133–147.
- de Cheveigné A (1993) Separation of concurrent harmonic sounds: fundamental frequency estimation and a time-domain cancellation model of auditory processing. *J Acoust Soc Am* 93:3271–3290.
- de Mare G (1948) Investigations into the functions of the auditory apparatus in perception deafness. *Acta Otolaryngol Suppl* 74:107–116.
- Evans EF (1978) Place and time coding of frequency in the peripheral auditory system: some physiological pros and cons. *Audiology* 17:369–420.
- Florentine M, Houtsma AJM (1983) Tuning curves and pitch matches in a listener with a unilateral, low-frequency hearing loss. *J Acoust Soc Am* 73:961–965.
- Fourcin AJ, Rosen SM, Moore BCJ, Douek EE, Clark GP, Dodson H, Bannister LH (1979) External electrical stimulation of the cochlea: clinical, psychophysical, speech-perceptual and histological findings. *Br J Audiol* 13:85–107.
- Freyman RL, Nelson DA (1986) Frequency discrimination as a function of tonal duration and excitation-pattern slopes in normal and hearing-impaired listeners. *J Acoust Soc Am* 79:1034–1044.
- Freyman RL, Nelson DA (1987) Frequency discrimination of short- versus long-duration tones by normal and hearing-impaired listeners. *J Speech Hear Res* 30:28–36.
- Freyman RL, Nelson DA (1991) Frequency discrimination as a function of signal frequency and level in normal-hearing and hearing-impaired listeners. *J Speech Hear Res* 34:1371–1386.
- Gaeth J, Norris T (1965) Diplacusis in unilateral high frequency hearing losses. *J Speech Hear Res* 8:63–75.

- Gengel RW (1973) Temporal effects on frequency discrimination by hearing-impaired listeners. *J Acoust Soc Am* 54:11–15.
- Geurts L, Wouters J (2001) Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants. *J Acoust Soc Am* 109:713–726.
- Glasberg BR, Moore BCJ (1986) Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments. *J Acoust Soc Am* 79:1020–1033.
- Glasberg BR, Moore BCJ (1990) Derivation of auditory filter shapes from notched-noise data. *Hear Res* 47:103–138.
- Goldstein JL, Srulovicz P (1977) Auditory-nerve spike intervals as an adequate basis for aural frequency measurement. In: Evans EF, Wilson JP (eds), *Psychophysics and Physiology of Hearing*. London: Academic Press, pp. 337–346.
- Grant KW (1987) Frequency modulation detection by normally hearing and profoundly hearing-impaired listeners. *J Speech Hear Res* 30:558–563.
- Grant KW, Ardell LH, Kuhl PK, Sparks DW (1985) The contributions of fundamental frequency, amplitude envelope, and voicing duration cues to speechreading in normal-hearing subjects. *J Acoust Soc Am* 77:671–677.
- Greenwood DD (1990) A cochlear frequency-position function for several species—29 years later. *J Acoust Soc Am* 87:2592–2605.
- Hall JW, Wood EJ (1984) Stimulus duration and frequency discrimination for normal-hearing and hearing-impaired subjects. *J Speech Hear Res* 27:252–256.
- Heinz MG, Colburn HS, Carney LH (2001) Evaluating auditory performance limits: I. One-parameter discrimination using a computational model for the auditory nerve. *Neur Comput* 13:2273–2316.
- Hoekstra A, Ritsma RJ (1977) Perceptive hearing loss and frequency selectivity. In: Evans EF, Wilson JP (eds), *Psychophysics and Physiology of Hearing*. London: Academic, pp. 263–271.
- Houtsma AJM, Goldstein JL (1972) The central origin of the pitch of pure tones: evidence from musical interval recognition. *J Acoust Soc Am* 51:520–529.
- Houtsma AJM, Smurzynski J (1990) Pitch identification and discrimination for complex tones with many harmonics. *J Acoust Soc Am* 87:304–310.
- Huss M, Moore BCJ, Baer T, Glasberg BR (2001) Perception of pure tones by listeners with and without a ‘dead region.’ *Br J Audiol* 35:149–150.
- Huss M, Moore BCJ (2005a) Dead regions and noisiness of pure tones. *Int J Audiol* (in press).
- Huss M, Moore BCJ (2005b) Dead regions and pitch perception. *J Acoust Soc Am* (in press).
- Ketten DR, Vannier MW, Skinner MW, Gates GA, Wang G, Neely JG (1998) In vivo measures of cochlear length and insertion depth of nucleus cochlear implant electrode arrays. *Ann Otol Rhinol Laryngol* 107:1–16.
- Kim DO, Molnar CE, Matthews JW (1980) Cochlear mechanics: nonlinear behaviour in two-tone responses as reflected in cochlear-nerve-fibre responses and in ear-canal sound pressure. *J Acoust Soc Am* 67:1704–1721.
- Lacher-Fougère S, Demany L (1998) Modulation detection by normal and hearing-impaired listeners. *Audiology* 37:109–121.
- Loeb GE, White MW, Merzenich MM (1983) Spatial cross correlation: a proposed mechanism for acoustic pitch perception. *Biol Cybern* 47:149–163.
- Long CJ, Carlyon RP, McKay CM, Vanat Z (2002) Temporal pitch perception: examination of first-order intervals. *Int J Audiol* 41:249.



- McDermott HJ, McKay CM (1994) Pitch ranking with nonsimultaneous dual-electrode electrical stimulation of the cochlea. *J Acoust Soc Am* 96:155–162.
- McDermott HJ, McKay CM (1997) Musical pitch perception with electrical stimulation of the cochlea. *J Acoust Soc Am* 101:1622–1631.
- McDermott HJ, McKay CM, Vandali AE (1992) A new portable sound processor for the University of Melbourne/Nucleus Limited multielectrode cochlear implant. *J Acoust Soc Am* 91:3367–3371.
- McDermott HJ, Lech M, Kornblum MS, Irvine DRF (1998) Loudness perception and frequency discrimination in subjects with steeply sloping hearing loss; possible correlates of neural plasticity. *J Acoust Soc Am* 104:2314–2325.
- McKay CM, McDermott HJ, Clark GM (1994) Pitch percepts associated with amplitude-modulated current pulse trains in cochlear implantees. *J Acoust Soc Am* 96:2664–2673.
- McKay CM, McDermott HJ, Clark GM (1995) Pitch matching of amplitude modulated current pulse trains by cochlear implantees: the effect of modulation depth. *J Acoust Soc Am* 97:1777–1785.
- McKay CM, O'Brien A, James CJ (1999) Effect of current level on electrode discrimination in electrical stimulation. *Hear Res* 136:159–164.
- McKay CM, McDermott HJ, Carlyon RP (2000) Place and temporal cues in pitch perception: are they truly independent? *Acoust Res Lett Online* (<http://ojpsaiporg/ARLO/tophtml>) 1:25–30.
- Meddis R, Hewitt M (1988) A computational model of low pitch judgement. In: Duifhuis H, Horst JW, Wit HP (eds), *Basic Issues in Hearing*. London: Academic Press, pp. 148–153.
- Meddis R, Hewitt M (1991) Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification. *J Acoust Soc Am* 89:2866–2882.
- Meddis R, O'Mard L (1997) A unitary model of pitch perception. *J Acoust Soc Am* 102:1811–1820.
- Micheyl C, Moore BCJ, Carlyon RP (1998) The role of excitation-pattern cues and temporal cues in the frequency and modulation-rate discrimination of amplitude-modulated tones. *J Acoust Soc Am* 104:1039–1050.
- Miller RL, Calhoun BM, Young ED (1999) Discriminability of vowel representations in cat auditory-nerve fibers after acoustic trauma. *J Acoust Soc Am* 105:311–325.
- Moore BCJ (1973a) Frequency difference limens for narrow bands of noise. *J Acoust Soc Am* 54:888–896.
- Moore BCJ (1973b) Frequency difference limens for short-duration tones. *J Acoust Soc Am* 54:610–619.
- Moore BCJ (1974) Relation between the critical bandwidth and the frequency-difference limen. *J Acoust Soc Am* 55:359.
- Moore BCJ (1977) Effects of relative phase of the components on the pitch of three-component complex tones. In: Evans EF, Wilson JP (eds), *Psychophysics and Physiology of Hearing*. London: Academic Press, pp. 349–358.
- Moore BCJ (1998) *Cochlear Hearing Loss*. London: Whurr.
- Moore BCJ (2001) Dead regions in the cochlea: diagnosis, perceptual consequences, and implications for the fitting of hearing aids. *Trends Amplif* 5:1–34.
- Moore BCJ (2003) *An Introduction to the Psychology of Hearing*, 5th ed. San Diego: Academic Press.
- Moore BCJ, Alcántara JI (2001) The use of psychophysical tuning curves to explore dead regions in the cochlea. *Ear Hear* 22:268–278.

- Moore BCJ, Glasberg BR (1986) The relationship between frequency selectivity and frequency discrimination for subjects with unilateral and bilateral cochlear impairments. In: Moore BCJ, Patterson RD (eds), *Auditory Frequency Selectivity*. New York: Plenum Press, pp. 407–414.
- Moore BCJ, Glasberg BR (1988a) Effects of the relative phase of the components on the pitch discrimination of complex tones by subjects with unilateral and bilateral cochlear impairments. In: Duifhuis H, Wit H, Horst J (eds), *Basic Issues in Hearing*. London: Academic Press, pp. 421–430.
- Moore BCJ, Glasberg BR (1988b) Pitch perception and phase sensitivity for subjects with unilateral and bilateral cochlear hearing impairments. In: Quaranta A (ed), *Clinical Audiology*. Bari, Italy: Laterza, pp. 104–109.
- Moore BCJ, Glasberg BR (1990) Frequency selectivity in subjects with cochlear loss and its effects on pitch discrimination and phase sensitivity. In: Grandori F, Cianfrone G, Kemp DT (eds), *Advances in Audiology*. Basel: Karger, pp. 187–200.
- Moore BCJ, Glasberg BR (1997) A model of loudness perception applied to cochlear hearing loss. *Audit Neurosci* 3:289–311.
- Moore BCJ, Moore GA (2003) Discrimination of the fundamental frequency of complex tones with fixed and shifting spectral envelopes by normally hearing and hearing-impaired subjects. *Hear Res* 182:153–163.
- Moore BCJ, Peters RW (1992) Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity. *J Acoust Soc Am* 91:2881–2893.
- Moore BCJ, Sek A (1994) Effects of carrier frequency and background noise on the detection of mixed modulation. *J Acoust Soc Am* 96:741–751.
- Moore BCJ, Sek A (1995) Effects of carrier frequency, modulation rate and modulation waveform on the detection of modulation and the discrimination of modulation type (AM vs FM). *J Acoust Soc Am* 97:2468–2478.
- Moore BCJ, Sek A (1996) Detection of frequency modulation at low modulation rates: evidence for a mechanism based on phase locking. *J Acoust Soc Am* 100:2320–2331.
- Moore BCJ, Skrodzka E (2002) Detection of frequency modulation by hearing-impaired listeners: effects of carrier frequency, modulation rate, and added amplitude modulation. *J Acoust Soc Am* 111:327–335.
- Moore BCJ, Glasberg BR, Peters RW (1985a) Relative dominance of individual partials in determining the pitch of complex tones. *J Acoust Soc Am* 77:1853–1860.
- Moore BCJ, Laurence RF, Wright D (1985b) Improvements in speech intelligibility in quiet and in noise produced by two-channel compression hearing aids. *Br J Audiol* 19:175–187.
- Moore BCJ, Wojtczak M, Vickers DA (1996) Effect of loudness recruitment on the perception of amplitude modulation. *J Acoust Soc Am* 100:481–489.
- Moore BCJ, Huss M, Vickers DA, Glasberg BR, Alcántara JI (2000) A test for the diagnosis of dead regions in the cochlea. *Br J Audiol* 34:205–224.
- Murray N, Byrne D (1986) Performance of hearing-impaired and normal hearing listeners with various high-frequency cut-offs in hearing aids. *Aust J Audiol* 8:21–28.
- Nelson PB, Jin S-H (2002) Understanding speech in single-talker interference: normal-hearing listeners and cochlear implant users. *J Acoust Soc Am* 111:2429.
- Nelson DA, van Tasell DJ, Schroder AC, Soli S, Levine S (1995) Electrode ranking of “place pitch” and speech recognition in electrical hearing. *J Acoust Soc Am* 98:1987–1999.

- Patterson RD (1976) Auditory filter shapes derived with noise stimuli. *J Acoust Soc Am* 59:640–654.
- Patterson RD (1987a) A pulse ribbon model of monaural phase perception. *J Acoust Soc Am* 82:1560–1586.
- Patterson RD (1987b) A pulse ribbon model of peripheral auditory processing. In: Yost WA, Watson CS (eds), *Auditory Processing of Complex Sounds*. Hillsdale, NJ: Erlbaum, pp. 167–179.
- Patterson RD, Wightman FL (1976) Residue pitch as a function of component spacing. *J Acoust Soc Am* 59:1450–1459.
- Patterson RD, Allerhand MH, Giguère C (1995) Time-domain modeling of peripheral auditory processing: a modular architecture and a software platform. *J Acoust Soc Am* 98:1890–1894.
- Pfingst BE, Holloway LA, Poopat N, Subramanya AR, Warren MF, Zwolan TA (1994) Effects of stimulus level on nonspectral frequency discrimination by human subjects. *Hear Res* 78:197–209.
- Pick G, Evans EF, Wilson JP (1977) Frequency resolution in patients with hearing loss of cochlear origin. In: Evans EF, Wilson JP (eds), *Psychophysics and Physiology of Hearing*. London: Academic Press, pp. 273–281.
- Pijl S, Schwarz DWF (1995) Melody recognition and musical interval perception by deaf subjects stimulated with electrical pulse trains through single cochlear implant electrodes. *J Acoust Soc Am* 98:886–895.
- Plack CJ, Carlyon RP (1994) The detection of differences in the depth of frequency modulation. *J Acoust Soc Am* 96:115–125.
- Plack CJ, Carlyon RP (1995) Differences in frequency modulation detection and fundamental frequency discrimination between complex tones consisting of resolved and unresolved harmonics. *J Acoust Soc Am* 98:1355–1364.
- Plack CJ, White LJ (2000) Pitch matches between unresolved complex tones differing by a single interpulse interval. *J Acoust Soc Am* 108:696–705.
- Plomp R (1967) Pitch of complex tones. *J Acoust Soc Am* 41:1526–1533.
- Plomp R, Steeneken HJM (1973) Place dependence of timbre in reverberant sound fields. *Acustica* 28:50–59.
- Risberg A (1974) The importance of prosodic elements for the lipreader. In: Nielson HB, Klamp E (eds), *Visual and Audio-visual Perception of Speech*. Stockholm: Almquist and Wiksell, pp. 153–164.
- Ritsma RJ (1963) On pitch discrimination of residue tones. *Int Audiol* 2:34–37.
- Ritsma RJ, Engel FL (1964) Pitch of frequency modulated signals. *J Acoust Soc Am* 36:1637–1655.
- Rosen S (1986) Monaural phase sensitivity: frequency selectivity and temporal processes. In: Moore BCJ, Patterson RD (eds), *Auditory Frequency Selectivity*. New York: Plenum Press, pp. 419–428.
- Rosen S (1987) Phase and the hearing impaired. In: Schouten MEH (ed), *The Psychophysics of Speech Perception*. Dordrecht: Martinus Nijhoff, pp. 481–488.
- Rosen S, Fourcin A (1986) Frequency selectivity and the perception of speech. In: Moore BCJ (ed) *Frequency Selectivity in Hearing*. London: Academic Press, pp. 373–487.
- Rosen SM, Fourcin AJ, Moore BCJ (1981) Voice pitch as an aid to lipreading. *Nature* 291:150–152.
- Ruggero MA (1994) Cochlear delays and traveling waves: comments on ‘Experimental look at cochlear mechanics.’ *Audiology* 33:131–142.

- Ruggero MA, Rich NC (1991) Furosemide alters organ of Corti mechanics: evidence for feedback of outer hair cells upon the basilar membrane. *J Neurosci* 11:1057–1067.
- Ruggero MA, Rich NC, Robles L, Recio A (1996) The effects of acoustic trauma, other cochlea injury and death on basilar membrane responses to sound. In: Axelsson A, Borchgrevink H, Hamernik RP, Hellstrom PA, Henderson D, Salvi RJ (eds), *Scientific Basis of Noise-Induced Hearing Loss*. Stuttgart: Thieme, pp. 23–35.
- Saberi K, Hafter ER (1995) A common neural code for frequency- and amplitude-modulated sounds. *Nature* 374:537–539.
- Schoeny Z, Carhart R (1971) Effects of unilateral Ménière's disease on masking level differences. *J Acoust Soc Am* 50:1143–1150.
- Schouten JF (1940) The residue and the mechanism of hearing. *Proc Konink Akad Wetenschap* 43:991–999.
- Sek A, Moore BCJ (1995) Frequency discrimination as a function of frequency, measured in several ways. *J Acoust Soc Am* 97:2479–2486.
- Sellick PM, Patuzzi R, Johnstone BM (1982) Measurement of basilar membrane motion in the guinea pig using the Mössbauer technique. *J Acoust Soc Am* 72:131–141.
- Shackleton TM, Carlyon RP (1994) The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. *J Acoust Soc Am* 95:3529–3540.
- Shamma SA (1985) Speech processing in the auditory system II: Lateral inhibition and the central processing of speech evoked activity in the auditory nerve. *J Acoust Soc Am* 78:1622–1632.
- Shamma S, Klein D (2000) The case of the missing pitch templates: how harmonic templates emerge in the early auditory system. *J Acoust Soc Am* 107:2631–2644.
- Shannon RV (1983) Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics. *Hear Res* 11:157–189.
- Shepherd RK, Javel E (1997) Electric stimulation of the auditory nerve. I. Correlation of physiological responses with cochlear status. *Hear Res* 108:112–144.
- Simon HJ, Yund EW (1993) Frequency discrimination in listeners with sensorineural hearing loss. *Ear Hear* 14:190–199.
- Simpson AM, Moore BCJ, Glasberg BR (1990) Spectral enhancement to improve the intelligibility of speech in noise for hearing-impaired listeners. *Acta Otolaryngol Suppl* 469:101–107.
- Skinner MW, Clark GM, Whitford LA, et al. (1994) Evaluation of a new Spectral Peak coding strategy for the Nucleus 22 channel cochlear implant system. *Am J Otol* 15: 15–27.
- Srulovicz P, Goldstein JL (1983) A central spectrum model: a synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum. *J Acoust Soc Am* 73:1266–1276.
- Summers V, Leek MR (1998) F0 processing and the separation of competing speech signals by listeners with normal hearing and with hearing loss. *J Speech Lang Hear Res* 41:1294–1306.
- Terhardt E (1974) Pitch of pure tones: its relation to intensity. In: Zwicker E, Terhardt E (eds), *Facts and Models in Hearing*. Berlin: Springer-Verlag, pp. 350–357.
- Thai-Van H, Micheyl C, Moore BCJ, Collet L (2003) Enhanced frequency discrimination near the hearing loss cutoff: a consequence of central auditory plasticity induced by cochlear damage? *Brain* 126:2235–2245.
- Thornton AR, Abbas PJ (1980) Low-frequency hearing loss: perception of filtered speech, psychophysical tuning curves, and masking. *J Acoust Soc Am* 67:638–643.

- Tong YC, Clark GM (1985) Absolute identification of electric pulse rates and electrode positions by cochlear implant listeners. *J Acoust Soc Am* 77:1881–1888.
- Tong YC, Blamey PJ, Dowell RC, Clark GM (1983) Psychophysical studies evaluating the feasibility of a speech processing strategy for a multiple-channel cochlear implant. *J Acoust Soc Am* 74:73–80.
- Townshend B, Cotter N, von Compernelle D, White RL (1987) Pitch perception by cochlear implant subjects. *J Acoust Soc Am* 82:106–115.
- Turner CW, Burns EM, Nelson DA (1983) Pure tone pitch perception and low-frequency hearing loss. *J Acoust Soc Am* 73:966–975.
- Tyler RS, Wood EJ, Fernandes MA (1983) Frequency resolution and discrimination of constant and dynamic tones in normal and hearing-impaired listeners. *J Acoust Soc Am* 74:1190–1199.
- van den Honert C, Stypulkowski PH (1987) Temporal response patterns of single auditory-nerve fibers elicited by periodic electrical stimuli. *Hear Res* 29:207–222.
- van Hoesel RJM, Clark GM (1997) Psychophysical studies with two binaural cochlear implant subjects. *J Acoust Soc Am* 102:495–507.
- Verschuure J, van Meeteren AA (1975) The effect of intensity on pitch. *Acustica* 32:33–44.
- Villchur E (1973) Signal processing to improve speech intelligibility in perceptive deafness. *J Acoust Soc Am* 53:1646–1657.
- Wakefield GH, Nelson DA (1985) Extension of a temporal model of frequency discrimination: intensity effects in normal and hearing-impaired listeners. *J Acoust Soc Am* 77:613–619.
- Webster JC, Schubert ED (1954) Pitch shifts accompanying certain auditory threshold shifts. *J Acoust Soc Am* 26:754–760.
- Wilson BS, Finley CC, Lawson DT, Wolford RD, Eddington DK, Rabinowitz WM (1991) Better speech recognition with cochlear implants. *Nature* 352:236–238.
- Wilson B, Zerbi M, Finley C, Lawson D, van den Honert C (1997) Speech processors for auditory prostheses (Eighth Quarterly Progress Report). NIH.
- Woolf NK, Ryan AF, Bone RC (1981) Neural phase-locking properties in the absence of outer hair cells. *Hear Res* 4:335–346.
- Zeng F-G (2002) Temporal pitch in electric hearing. *Hear Res* 174:101–106.
- Zeng FG, Fu QJ, Morse R (2000) Human hearing enhanced by noise. *Brain Res* 869:251–255.
- Zeng F-G, Popper AN, Fay RR (2004) *Auditory Prostheses*. New York: Springer-Verlag.
- Zurek PM, Formby C (1981) Frequency-discrimination ability of hearing-impaired listeners. *J Speech Hear Res* 24:108–112.
- Zwicker E (1956) Die elementaren Grundlagen zur Bestimmung der Informationskapazität des Gehörs. *Acustica* 6:356–381.
- Zwicker E, Fastl H (1990) *Psychoacoustics—Facts and Models*. Berlin: Springer-Verlag.
- Zwolan TA, Collins LM, Wakefield GH (1997) Electrode discrimination and speech recognition in postlingually deafened adult cochlear implant subjects. *J Acoust Soc Am* 102:3673–3685.

# Pitch and Auditory Grouping

CHRISTOPHER J. DARWIN

## 1. Introduction

How often do you hear a single sound by itself? Only when doing psychoacoustic experiments in a soundproof booth! In our everyday environment, there is almost always more than one sound present. Sounds that have a pitch—speech, musical notes, bird song—are usually encountered in the context of other similar sounds—in the pub, at a concert, or in the woods.

Despite this rather obvious fact, almost all the research in pitch perception over the last 150 years has been aimed at understanding how humans perceive the pitch of a single pure or complex tone presented alone. Why? One could argue that the simple problem of how humans perceive the pitch of a single sound should be understood first, before attempting the undoubtedly more difficult problem of perceiving the pitches of multiple simultaneous sounds. But that strategy could be misleading, with theories developing that, although adequate for single sound sources, could generalize only with difficulty to multiple sources. One reason that they might fail in this way is by assuming that all the sound present at a particular time is relevant to working out the pitch of just one of the sounds.

Understanding how humans perceive the pitch of each of a number of simultaneous sounds is part of the more general problem of how we perceive all the attributes of simultaneous sounds: their separate locations and timbres as well as their pitches and how they change over time (for a general review of auditory grouping see Darwin and Carlyon 1995). How do we confine the decision making about a single sound source to only the components that originate from that sound? A general approach to the problem of how we segregate a sound mixture into groups that correspond to different sound sources has been described by Albert Bregman (Bregman 1990) in his influential book, *Auditory Scene Analysis*. Bregman distinguishes two different strategies: primitive and schema-based segregation. Bregman's primitive grouping mechanisms use general constraints on sound sources and are described by him as preattentive and

innate, whereas schema-based constraints invoke learned knowledge about particular sounds.

One example of a primitive constraint is onset time. When an object is struck, or otherwise made to vibrate, the different frequencies that it produces as it vibrates start at roughly the same time. So a useful heuristic for grouping sounds from a common sound source is to treat those that start at the same time as belonging together. As composers well know, accurate synchronization of two notes can mislead the brain into interpreting them as a single note with a new, composite timbre. An example from the classical music literature is the blending of the oboe and clarinet in unison at the opening of Schubert's *Unfinished Symphony* (Broadbent and Ladefoged 1957). With skilful, well-synchronized players it is impossible to tell what those two instruments are, or even that there are two instruments at all. With less skilled, poorly synchronized performers, the two instruments maintain their individual identities and the novel composite timbre is lost.

Another example of a primitive constraint is harmonicity—the harmonic sequence of frequency components that makes up a periodic sound. Two sounds on the same fundamental frequency ( $F_0$ ) are harder to separate than two on different  $F_0$ s—the oboe and clarinet in the Schubert example play in unison. More generally, two sounds whose combined frequency components make up a single harmonic series will tend to fuse into a single auditory object. A musical example is the “12th” or “octave quint” organ stop which produces a tone at an interval of a 12th (octave plus 5th) above the note played on the keyboard. When used in addition to a traditional stop, this stop adds a tone at triple the  $F_0$  of the original note. The harmonics of the higher note thus coincide with every third harmonic of the original and so simply modify the timbre without producing the impression of a separate pitch. Were the “12th” to be sufficiently out of tune (or out of synchrony) with the original played note, it would stand out as a separate note at the higher pitch and both notes would maintain their original timbre.

In this chapter we will look at two related topics. The first is the evidence for the use of primitive grouping constraints such as harmonicity and onset time in the perception of pitch. The second is the use of harmonicity as a primitive grouping constraint to help us to establish the timbre or the location of the constituent sound sources of a mixture.

## 2 Grouping in Pitch Perception

It is well established that the pitch of a broadband complex sound is determined predominantly by the frequencies of the low-numbered, resolved harmonics (see Plack and Oxenham, Chapter 2). Goldstein's (1973) influential model of pitch perception finds the best-fitting harmonic series to the sequence of resolved harmonics. This model explains a very substantial body of data on the percep-

tion of the pitch of single complex sounds, that is, sounds that can be appropriately matched by a single harmonic series. However, the model does not address the important problem of how to estimate pitch when more than one periodic sound is present. Simply finding the best-fitting harmonic series to all the frequencies that are resolved from a mixture of two complex tones will give a single answer that does not correspond to the perceptual reality of two distinct pitches. Although this problem is strikingly obvious when two different-pitched sounds are present, it also applies in practical situations in which one is trying to estimate the pitch of, say, speech. In speech, the pitch and the timbre of the sound may change rapidly, giving a sound that is not truly periodic; consequently the frequency estimates of the individual harmonics may be rather variable.

## 2.1 Harmonicity

A sensible rule-of-thumb that would provide some leverage on the problem of which frequency components to take into account when estimating the pitch of a complex is to consider only those frequency components that lie sufficiently close to a harmonic frequency of the pitch being considered. This principle underlies the “harmonic sieve” (Duifhuis et al. 1982), which was programmed as a front end to an implementation of Goldstein’s pitch mechanism for estimating the pitch of natural speech. The harmonic sieve effectively excludes from the calculation of pitch any component whose frequency lies more than some fixed percentage from a harmonic of  $F_0$ .

This heuristic is also used by human listeners; the tolerance that they give to individual harmonics has been addressed experimentally by Moore and his colleagues (Moore et al. 1985a). They mistuned one harmonic of a 12-harmonic complex, and measured the consequent shift in the pitch of the complex. For small mistunings (less than about 3%) the pitch shift was a roughly linear function of the mistuning, but for larger mistunings the pitch shift of the complex *decreased*, approaching zero by about 8% mistuning. Their results show that the harmonic sieve does not work on an “all-or-none” basis; rather, a harmonic makes progressively less contribution to the pitch of the complex as its mistuning increases from 3% to beyond 8%.

Moore’s results have been extended to a larger number of values of mistuning (Darwin 1992; Darwin and Ciocca 1992). Some of these data are shown in Figure 8.1. They can be well fitted by assuming that the contribution that a particular harmonic makes to the pitch of a complex sound varies according to a Gaussian function of the amount of mistuning, with the width of the Gaussian (parameter  $s$  in the figure) being around 3%. Parameter  $k$  in the figure is a measure of how much of a contribution overall the mistuned harmonic makes to the pitch. Moore’s original data showed that the low-numbered harmonics make more of a contribution than the higher-numbered, but there is considerable variability across listeners in the relative importance of the different low-numbered harmonics (Moore et al. 1985a).



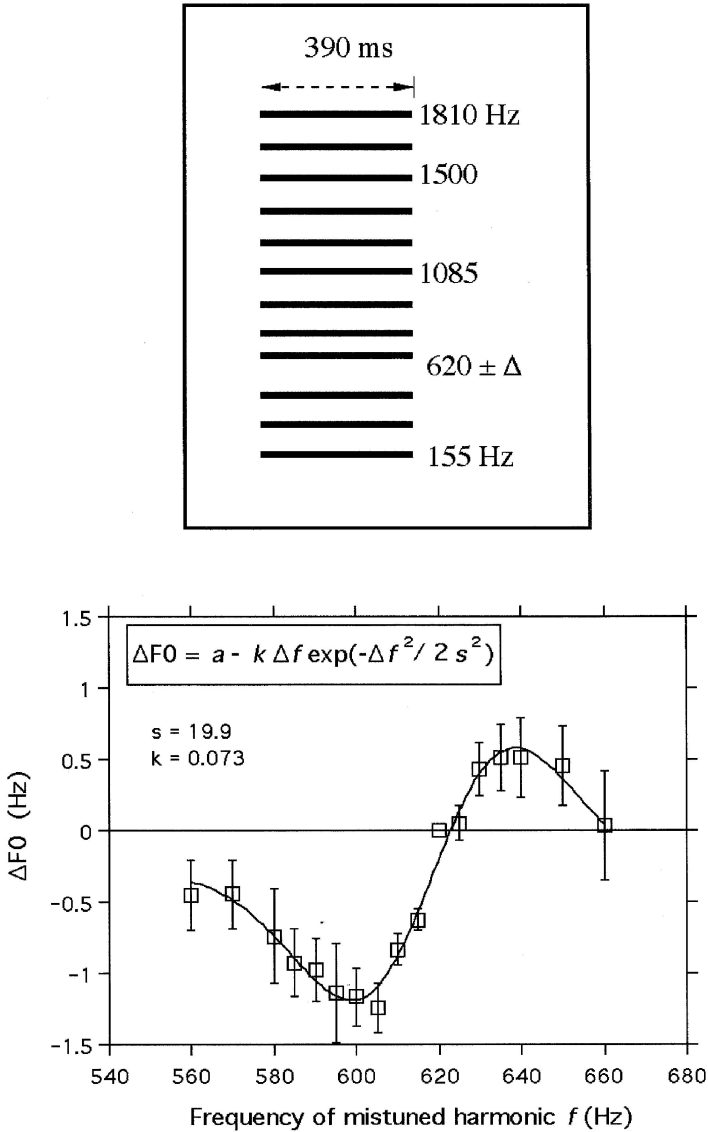


FIGURE 8.1. Matched pitch shifts (from 155 Hz) produced by mistuning the 4th harmonic of a 12-harmonic complex with a fundamental of 155 Hz. The fitted curve assumes that the contribution that a progressively mistuned harmonic makes to the perceived pitch varies according to a Gaussian function with a standard deviation of  $s$ .

The figure of roughly 8% for the tolerance of the human “harmonic sieve” fits well with the tolerance used by Duifhuis et al. (1982) in their program for extracting pitch from natural speech. It seems likely therefore that some such selection of frequency components with harmonically plausible frequencies would be a necessary front end to human pitch perception if it operated in a way broadly similar to Goldstein’s theory. But Goldstein’s is not the only candidate for a theory of pitch perception. Could the results that we have just presented be predicted by an autocorrelation theory?

In principle they could. A strictly periodic sound will produce a clear peak in, for example, a summary autocorrelation function (SACF) (Meddis and O’Mard 1997), or in a histogram of first-order spike intervals (Moore 1987). Since a mistuned harmonic is strictly periodic at a slightly different period from that of the rest of the sound, it would by itself produce a peak at a slightly different period from that of the rest of the sound. For small mistunings, the peak of the complete sound would thus shift. For larger mistunings, a separate peak due to the mistuned harmonic would appear on the flank of the main peak, and the period of the main peak would then be determined primarily by the in-tune harmonics.

Although such shifting peaks provide a neat qualitative explanation of the effects of mistuning a harmonic, Meddis and O’Mard (1997, Figure 7) found a substantial quantitative discrepancy between the predictions of their autocorrelation model and the experimental data. The model predicted a tolerance that was about double that of the experimental data. So at least the Meddis and O’Mard version of an autocorrelation model requires some additional segregation of frequency components in order to give results that match those of human listeners.

In summary, both the Goldstein and the Meddis and O’Mard models of pitch perception require some preliminary sorting of frequency components before they can both match the performance of human listeners and also provide robust performance on natural signals. This sorting rejects from the calculation of pitch those components that deviate too far from a harmonic frequency of the pitch. Some such sorting mechanism is assumed by Beerends and Houtsma (1986) in their application of Goldstein’s theory to the results of their experiments on the perception of two simultaneous pitches, each generated by two harmonics. The Goldstein model provides good independent estimates of the two pitches provided that the processor knows that there are two pitches present, how to pair the two pairs of harmonics, and what the set of allowed F0s is.

Slightly mistuning a single harmonic of a complex not only produces a shift in the pitch of the complex, but it also, somewhat inconsistently, makes the mistuned harmonic stand out as a separate sound. The inconsistency is that the auditory system is treating the mistuned harmonic both as a separate sound—listeners can tell which harmonic is mistuned when the mistuning is only about 1% to 2% (Moore et al. 1985b)—and as contributing to the pitch of the complex. Much larger mistunings are required to prevent the mistuned harmonic contributing to the pitch of the complex than to make the mistuned component audible as a separate sound (with its own pitch). This is a simple example of a phe-

nomenon first reported in speech perception (where it was referred to as duplex perception; Liberman et al. 1981) in which the same component of a sound may make independent contributions to two different percepts (Bregman 1987). It is also an example of the important principle that the extent to which sounds group together depends on what property of the sound is being perceived (Hukin and Darwin 1995).

Although periodic sounds are harmonic, with their frequency components at integer multiples of  $F_0$ , it may be that the auditory system is sensitive to the linear spacing of components in a harmonic pattern rather than to its strict harmonicity (Roberts and Brunstrom 1998). Roberts and Brunstrom (2001) constructed complex sounds in which each member of a harmonic series was shifted in frequency by a constant amount. This manipulation maintains the equal spacing of the original series, but it is no longer harmonic. Their listeners could hear out a single component that deviated from a moderately shifted pattern almost as easily as one that deviated from the original unshifted, harmonic pattern. However, this pattern of results may still be explicable in terms of a model such as autocorrelation, which is based on the detection of periodicity (Roberts and Brunstrom 2001).

### 2.1.1 Joint or Disjoint Allocation

So far we have talked of grouping and segregation as if there were a set of discrete frequency components that were to be allocated uniquely to one sound source or another. But the truth is more complex. Because of the limited resolving power of the ear, frequency components from different sound sources will frequently fall within the same auditory filter. In the laboratory, or with expert tuning of instruments, they may even coincide in frequency. Does the pitch perception system take account of this possibility, or does it allocate the sound in an individual auditory filter only to one sound source?

For simultaneously presented sounds, there is some evidence that a particular frequency channel can contribute to the pitch of two simultaneous complex sounds. Beerends and Houtsma (1989) report that when a complex with frequencies 1067 and 1333 Hz is played together with one with frequencies 800 and 1000 Hz all to the same ear, then pitches of 200 and 267 Hz are reported even though the frequencies 1000 and 1067 Hz fall within the same auditory filter. However, these frequencies may be sufficiently separated (a semitone apart) that there is enough information in the phase-locked auditory response to these sounds in the skirts of their excitation pattern to enable the brain to determine their two frequencies.

A different approach to this problem (Darwin 1992) exploits the shift in pitch of a complex that results when a harmonic is mistuned. Listeners heard two complex tones on different  $F_0$ s. The  $F_0$ s were constructed so that the third harmonic of the higher-pitched tone differed in frequency from the fourth harmonic of the lower-pitched tone by 3%. These two frequency components were replaced in the mixture by a single tone whose frequency could be varied. When this adjustable tone is exactly at a harmonic frequency for one complex it would

be 3% mistuned in the other, and so capable of producing a maximum pitch shift in it. However, if the tone were disjointly allocated to the harmonic series for which it is in tune, it would produce no pitch shift in the complex for which it was mistuned. The results indicated that the mistuned harmonic is not disjointly allocated to only one harmonic series. The auditory system appears to make independent decisions on the pitches of the two complexes, so that even when the component is perfectly in tune with one complex, it still influences the pitch of the other. Such an outcome would be expected from both the harmonic sieve (with different harmonic sieves being applied independently to a mixture) and autocorrelation models discussed earlier.

## 2.2 *Onset Synchrony*

Another sensible rule of thumb for determining what frequency components come from what sound is to group together sounds that have similar onset times. Algorithms for extracting the pitches from sound mixtures perform better if some account is taken of the relative times at which different frequency components start (Denbigh and Zhao 1992). Another example comes from the art of musical composition. In polyphonic music (such as a fugue) the composer aims to maintain the integrity of each individual voice or instrument whereas in homophonic music (such as a chorale or a simply harmonized hymn tune) the aim is to provide an integrated perceptual texture. Huron (2001) identifies a difference in onset time as an important principle in ensuring the perceptual independence of parts, and as the most important factor distinguishing polyphonic from monophonic scoring.

Figure 8.2 shows a narrow-band spectrogram of a 3-s excerpt from one of Bach's Goldberg variations arranged for strings. At any one moment harmonics from three or four different instrumental pitches are present, but the frequency components that start together tend to be from a single instrument and are harmonically related.

For many sounds produced percussively, the onset times of the different components are very similar (although the subsequent growth and decay rates of different components may be very different). However, for periodic sounds produced by string and wind instruments (including the human voice), the onset times of different components can be spread over a 10th of a second or more. Indeed, the way in which the different components start is an important aspect of an instrument's timbre: instrument identification is worse for sounds that have the onset transient removed (Saldanha and Corso 1964).

Experimental evidence on how onset time influences pitch perception has established that the pitch of a complex is not influenced by individual frequency components that start substantially before the rest. The experiments that have shown such an effect of onset-time (Darwin and Ciocca 1992; Ciocca and Darwin 1993) have exploited the pitch shift produced by a mistuned harmonic that was described in the previous section.

When the 4th harmonic of a 12-harmonic periodic complex sound is mistuned

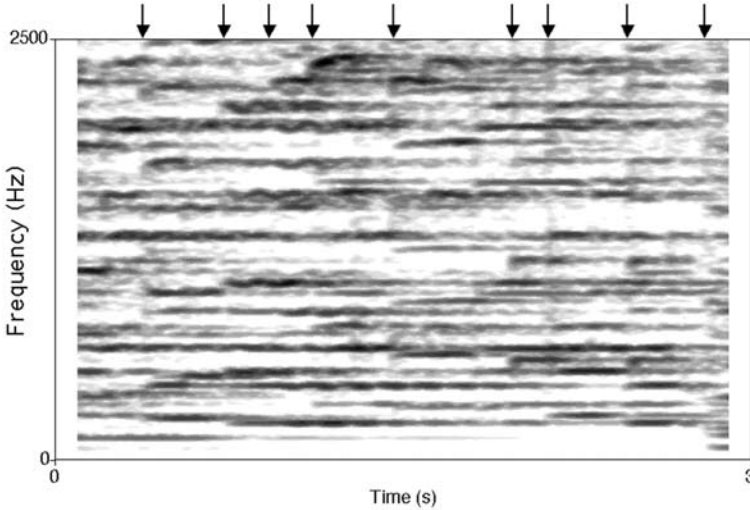


FIGURE 8.2. Narrow-band spectrogram of a 3-s excerpt from one of J.S. Bach's Goldberg Variations arranged for strings. At any one moment harmonics from up to four instrumental voices are present, but those frequency components that start together are usually from a single instrument and so are harmonically related. The times at which some of the notes start are indicated by *vertical arrows*.

by +3%, the pitch of the complex increases slightly. However, this change can be removed by allowing the mistuned harmonic to start earlier than the rest (left panel of Fig. 8.3). Surprisingly large amounts of onset asynchrony are needed to effect this removal. For a 90-ms complex, an onset asynchrony of around 150 ms is needed to remove the leading, mistuned harmonic from the calculation of pitch. This perceptual removal of the leading harmonic could have a rather simple explanation. Perhaps the auditory system's response to the harmonic has simply adapted during the lead time, so that by the time that the other components start, only an attenuated auditory representation of the leading harmonic is present. This explanation is unlikely to be the whole story.

The right panel of Figure 8.3 shows another complex added to the configuration shown in the left panel, which is synchronous with just the leading portion of the 640-Hz tone, and harmonically related to it ( $F_0$  of 213 Hz). With this configuration the effect of the onset asynchrony is much reduced—most of the pitch shift remains. Although the additional complex would have no influence on any adaptation that is occurring to the 640-Hz tone, it *is* effective at perceptually removing the leading part of the 640-Hz tone, thereby allowing the remainder of that tone to contribute to the pitch of the 155-Hz complex.

These experiments reject one style of model of pitch perception, which we might call the *bacon-slicer tendency*. In such a model, the output of the ear's spectral analysis of sound is cut into temporal slices, and the pitch of the sounds

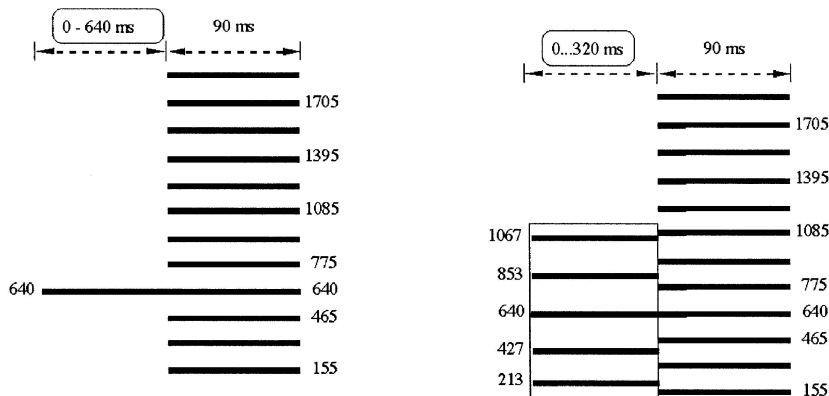


FIGURE 8.3. Stimuli used to demonstrate effect of onset-asynchrony on pitch perception. In the left-hand panel, all but one of the components are harmonics of a 155-Hz fundamental. The 640-Hz component is sharpened by 3% from the harmonic frequency of the fourth harmonic of 155 Hz (620 Hz). When this mistuned component is synchronous with the other harmonics the pitch of the complex is about 1 Hz sharp of 155 Hz, but as the 640-Hz component is given an increased onset time, the pitch of the complex returns to that of a periodic 155-Hz tone. In the right-hand panel, the leading portion of the mistuned component is grouped with a different harmonic complex, thereby destroying the effect of the onset asynchrony and increasing the pitch shift of the 155-Hz complex that now perceptually includes the continuation of the 640-Hz tone.

in each slice determined without regard for the past history (or future prospects) of the components within each slice. Each slice of spectral bacon is thus classified independently of the content of neighboring slices. Such a model would fail to parse frequency components into source-related groups on the basis of their differing time courses, and so would include all sufficiently harmonic components into the calculation of pitch. As the experiments on onset time have shown, the auditory system behaves more intelligently than this, and will discount a sufficiently harmonic component if it started a sufficiently long time before the other components in a complex, provided it is not itself temporally subdivided by other groupings.

The general principle operating here is what Bregman (1990) has termed the “Old plus New” heuristic. If a sound becomes suddenly more complex or more intense, the auditory system tries to interpret this change as a continuing old sound being joined by a new one. The “old,” leading tone is thus interpreted as a separate sound continuing into the “new,” later-starting components. The Old plus New interpretation here is strengthened by the continuity of the leading component; but similar Old plus New context effects have been shown in pitch perception where sounds are repeated but are not continuous (see Section 2.3).

Although this principle works well for the low-numbered resolved harmonics of a complex sound, it appears not to be applicable when we consider the perception of sounds that consist only of high-numbered unresolved harmonics.

The pitch of unresolved harmonics is carried by the repetition rate of the envelope of the sound regardless of the spectral region that the sound is in. This repetition rate persists after cochlear filtering. Its perception is probably achieved by timing the intervals between auditory nerve spikes that are phase locked either to maxima in the envelope, or to local maxima in the waveform that are close to envelope maxima (see Plack and Oxenham, Chapter 2). This mechanism is capable of giving at least a modest pitch sensation when only a single sound is present (Houtsma 1984). It is also likely to be able to work effectively when there is sufficiently little overlap in the frequency content of sounds with different periodicities. However, when two sounds that occupy *the same spectral region* have different periodicities, listeners find it impossible to hear two distinct pitches. Instead, the percept degenerates into a noisy crackle (Carlyon 1996a). The reason for this lack of perceptual clarity can be seen in Figure 8.4. The two top panels show the output of an auditory filter in response to each of two single complexes—with the higher pitch in the top panel. The bottom panel shows the output when the two sounds are mixed together. To the eye as well as to the ear the mixture is not readily decomposable into two periodicities. This result has implications for the information available to the periodicity detection mechanism. A simple autocorrelation model that uses all the information in the auditory nerve should show peaks corresponding to each of the two constituent periodicities (see also Kaernbach and Demany 1998).

Not only can listeners not hear the constituent pitches but they are also unable to use a difference in onset time between the two complex sounds in order to separate out the two constituent pitches (Carlyon 1996a,b). These observations set interesting limits to the effectiveness of Bregman's "Old plus New" heuristic. It may be that the auditory system can use this heuristic only to allocate to sound sources different proportions of energy in auditory filter channels (Darwin 1995; McAdams et al. 1998), and that it is unable to partition more abstract properties.

### 2.3 Context

If a complex tone consisting of two simultaneous components with frequencies  $f_1$  and  $f_2$  is embedded in a sequence of tones of frequency  $f_1$ , listeners will be torn between integrating the  $f_1$ - $f_2$  complex into a whole, and segregating it so that its  $f_1$  can become part of the surrounding sequence. In the latter case, the complex decomposes into an "old"  $f_1$  and a "new"  $f_2$  according to the "Old plus New" heuristic (Bregman and Pinker 1978). A similar decomposition occurs in pitch perception.

The upper panel of Figure 8.5 shows a complex sound with its 4th harmonic mistuned, preceded by four repetitions of this mistuned harmonic. Listeners matched the pitch of the complex as a function of the amount of mistuning of the 4th harmonic. When the complex was played by itself, the pitch of the complex shifted in a similar way to that found in previous experiments—with a maximum shift in pitch of about 1% at a mistuning of around 3% to 4%. However, when the complex was preceded by four tones at the same frequency

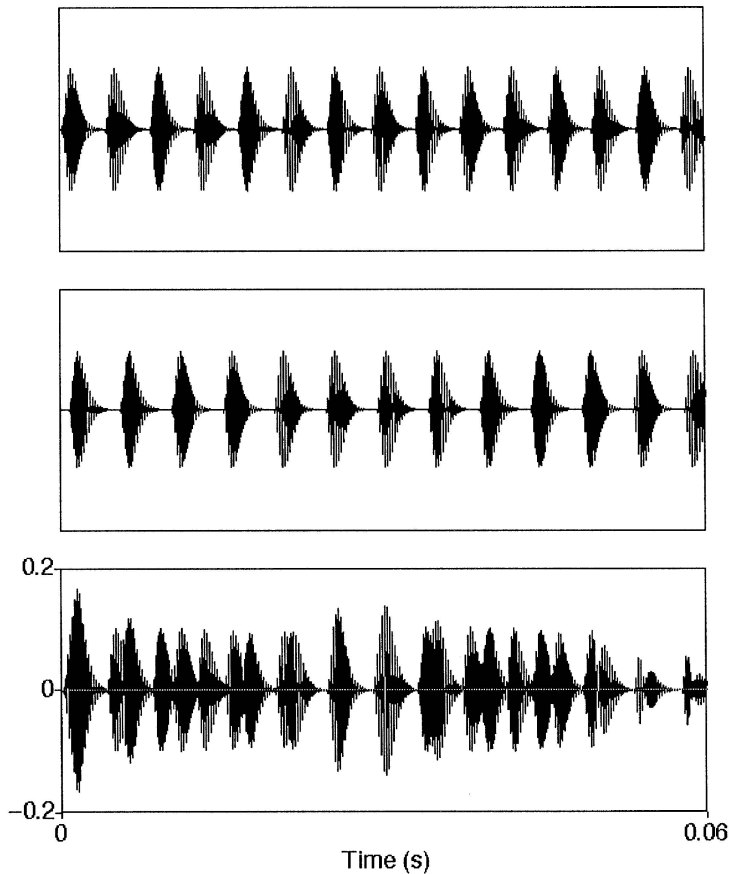


FIGURE 8.4. Each panel shows the output of an auditory filter centered at 4.5 kHz in response to complex tones with periodicities of (*top*) 243.6 Hz, (*middle*) 210 Hz, and (*bottom*) 210 Hz plus 243.6 Hz.

as the mistuned 4th harmonic, the pitch shift disappeared indicating that the mistuned harmonic had formed a perceptual stream with the preceding four similar tones, and removed it from the complex (Darwin et al. 1995).

## 2.4. Localization Cues

At first sight, one might think that a powerful heuristic would be to group together sounds that come from a common location. In some situations, for example, in sequential grouping (see Section 3), a common spatial direction does indeed lead to powerful grouping, but in the simultaneous grouping of harmonics the auditory system appears largely to ignore localization cues. Why it should behave in this way is an intriguing question.



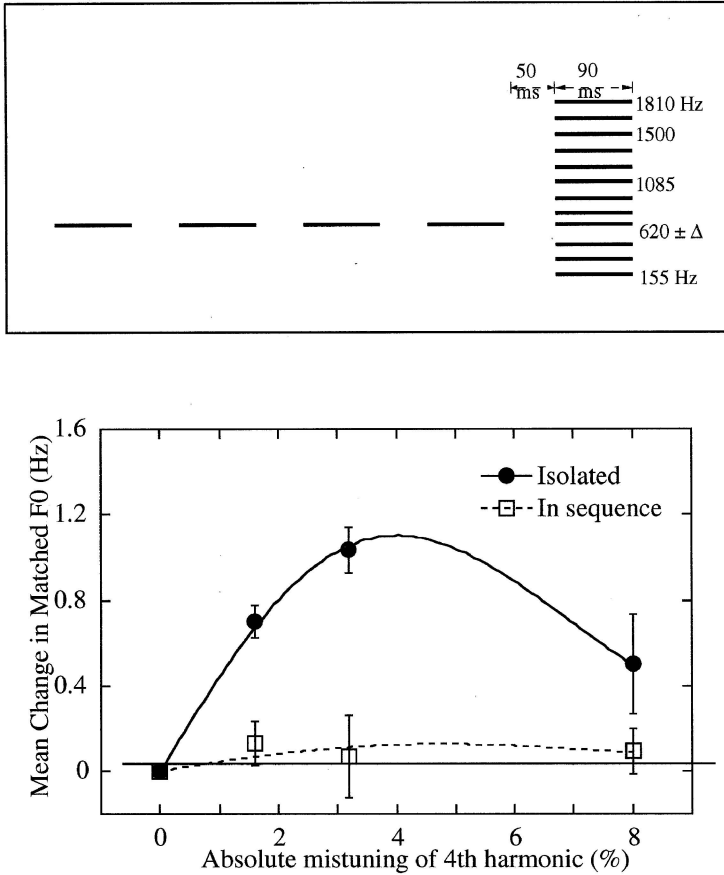


FIGURE 8.5. The *upper panel* shows the stimulus configuration used to demonstrate an effect of a repeating context on pitch perception. The 4th harmonic of the complex is mistuned, and the complex optionally preceded by four repetitions of the tones identical to the mistuned harmonic. The *lower panel* shows the results of pitch matches to the complex. The mistuned harmonic shifts the pitch of the complex heard in isolation, but not when it is preceded by the tone sequence.

The human auditory system uses two main cues to localize sound in the horizontal plane (or azimuth): interaural time difference (ITD) and interaural level difference (ILD).

ITDs arise because sound from a source that is to one side of the midline has further to travel to reach the opposite ear than to reach the one on the same side of the head. The maximum difference for an adult is a little more than half a millisecond. There are cells in the mammalian brainstem specialized for detecting these small time differences. ITDs provide unambiguous information predominantly for low spectral frequencies (below about 750 Hz). For complex

wide-band sounds such as speech, the ITDs of the low-frequency components provide the main localization cue (Wightman and Kistler 1992).

ILDs, on the other hand, arise from two different causes. First, for close sounds, the sound at the further ear is quieter simply because it has traveled further—the inverse-square law dictates that for every doubling of distance, a sound has its energy reduced to a quarter (a reduction of 6 dB). This change applies to all frequencies of sound, but becomes negligible (less than 1 dB) when sounds are further away than a couple of meters. Second, the head casts an acoustic shadow, so that sounds at the farther ear are less intense than those at the nearer ear. The shadow is darker for higher frequency sounds (around 20 dB at 4 kHz) and is negligible for frequencies less than a few hundred Hertz, but it does apply equally to sounds at all distances.

An easy and common procedure for presenting two sounds from different directions is to present them dichotically—with each sound presented to a different ear. In terms of natural cues, this form of presentation maximizes ILD (it is in principle infinite) but ITDs will be undefined, when, as is usual, the sounds on the two ears are composed of different frequencies. Even this extreme form of presentation has almost no effect on grouping in pitch perception. If two consecutive harmonics from two different fundamentals are presented simultaneously, listeners are no better at identifying the two fundamentals when the harmonics are appropriately segregated by ear than when each ear receives one harmonic from each fundamental (Beerends and Houtsma 1986). A similar conclusion can be drawn from data gathered by Darwin and Ciocca (1992), measuring the pitch shift produced by a single mistuned harmonic. Their data shown in Figure 8.1 are for the case where the mistuned component is presented to the same ear as the rest of the complex, but very similar results were obtained when the mistuned component was led to the opposite ear.

These experiments have all used dichotic (infinite ILD) presentation. Although similar experiments using ITDs have not been done, it is clear from other grouping experiments that ITDs are generally less effective than ILDs in promoting simultaneous segregation (Culling and Summerfield 1995).

Why should lateralization cues have such little effect on grouping in pitch perception, when harmonicity, onset time, and contextual effects such as repetition have marked effects? The problem is not unique to pitch perception, since grouping for timbre in speech perception shows remarkably little tendency to use ITDs (Culling and Summerfield 1995) for grouping (although here infinite ILDs do prove effective). The answer may be that in noisy and/or reverberant environments, localization cues in any one frequency channel are not sufficiently robust to provide the basis for consistent and reliable grouping decisions. Echoes and other sound sources can seriously disrupt lateralisation cues in individual frequency channels. Yet our percept of the location of a sound source is remarkably stable: you never hear the different frequency regions of a sound coming simultaneously from different locations. It may be that we decide what simultaneously present components make up a sound before we decide where that sound is. The simultaneous grouping might then be based both on low-

level grouping cues and also on schema-based cues; the low-level grouping cues would include harmonicity, onset time and temporal context, but would not include localization information (Woods and Colburn 1992; Darwin and Hukin 1999). Once the frequency composition of a sound source is determined, then its location could be calculated by pooling the localization cues from the component frequency channels. Provided that the grouping of individual frequencies into auditory objects was carried out effectively, pooling localization estimates across the frequency components that formed an object should lead to a stable percept of that object's position.

### 3. Using Harmonicity and Pitch in Grouping

In the second part of this chapter I will review work on the way that harmonicity is used to group simultaneous sounds together in the perception of the timbre or the direction of the sound source. I will also look at how harmonicity or pitch is used to group sounds together across time.

#### *3.1 Separating Simultaneous Sounds by F0*

##### 3.1.1 Timbre

When two periodic sound sources are active at the same time, it will usually be the case—except in a musical ensemble—that they have different F0s. The corresponding difference in harmonic structure provides an important cue for perceptual segregation of the two simultaneous sounds. Scheffers (1979, 1983) provided the original demonstration of the improvement in recognition that a difference in F0 between two sounds can provide. He played his listeners pairs of 220-ms duration simultaneous vowels (chosen from a set of eight Dutch vowels) that had been synthesized to be either on the same F0 or on different F0s. He found that listeners' correct identification of both vowels in a pair improved from about 40% correct to about 60% correct when the F0 difference between the vowel pairs increased from zero to one semitone. This basic result has been replicated for English (Assmann and Summerfield 1990; Culling and Darwin 1993), German (Zwicker 1984), and French (de Cheveigné et al. 1995) vowels, and a very similar result obtained for the recognition of pairs of five orchestral instruments—flute, Bb clarinet, cor anglais, French horn, and viola—from steady 1000-ms duration notes with natural onsets (Sandell and Darwin 1996). The consistent pattern of results from all these studies is that although identification is well above chance when the sounds have the same F0, it increases by about 20% as the F0 difference increases to one semitone, but then asymptotes for further F0 increases.

A difference in F0 gives a somewhat different pattern of results when simultaneous passages of fluent speech are used instead of isolated steady-state vowels. Brokx and Nootboom (1982) presented their listeners with individual target nonsense sentences against a background of continuous speech. Both the

target sentences and the background speech were manipulated using linear-predictive coding to give flat F0 contours at different values of F0. The words of the nonsense sentences became more intelligible as the difference in F0 between them and the background speech was increased up to three semitones. Brokx and Nootboom used only a single value larger than three semitones—twelve semitones, which gave performance that was close to that with no F0 difference. Why should performance for isolated vowels asymptote at one semitone, whereas performance for fluent speech increases out to at least three semitones?

The answer lies partly in the distinction between simultaneous and sequential grouping, and partly in the way that a difference in F0 allows simultaneous grouping to occur.

To successfully follow one voice in the presence of another the listener must solve two problems: first, to segregate the simultaneous components into groups that correspond to the different voices and second, to link together across time those groups that belong to the same voice. So, if at one time there are two groups of components A and B, and at a later time there are another two groups X and Y, then is X or Y the continuation of A? This problem is discussed in the following section on sequential grouping, but for the present we can note that continuity of the pitch of a voice is likely to contribute to the ease of following a particular voice. The second part of the answer is more complex. How does a difference in F0 help in simultaneous grouping?

There are two different ways in which a difference in F0 could help to improve the intelligibility of two simultaneous vowel sounds. The most obvious way, across-formant grouping, was originally suggested by Broadbent and Ladefoged (1957). Sounds in different spectral regions are grouped by virtue of a common harmonic series or periodicity. Consider the following simple example from speech. The upper panel of Figure 8.6 shows the spectra of two vowels /a/ on an F0 of 100 Hz and /i/ on an F0 of 140 Hz. The /i/ has its first two formants at 300 and 2500 Hz, and the /a/ has its first two formants at 440 and 800 Hz. In the region of a vowel's formant frequency, harmonics from that vowel have a higher amplitude than do those from the other vowel, and so would dominate the auditory representation of the mixture. So the first formant of /i/ will dominate the spectrum of the mixture in the region around 250 Hz and its second formant around 2000 Hz; similarly the first two formants of /a/ will dominate the spectrum from about 400 Hz through 1500 Hz. Within these regions the auditory representation of the sound will convey the harmonic structure or periodicity of the dominant vowel. Broadbent and Ladefoged proposed that the common harmonic structure in say the 300- and 2500-Hz regions might allow the auditory system to treat them as part of the same sound source, and as a different sound source from the intervening region. Some such process does occur in speech. For example, Darwin (1981) produced a four-formant syllable that in its entirety was heard as /ru/, but when the second formant was physically removed, was heard as /li/. The /li/ percept could also be obtained even when all four formants were physically present by putting the second formant on a

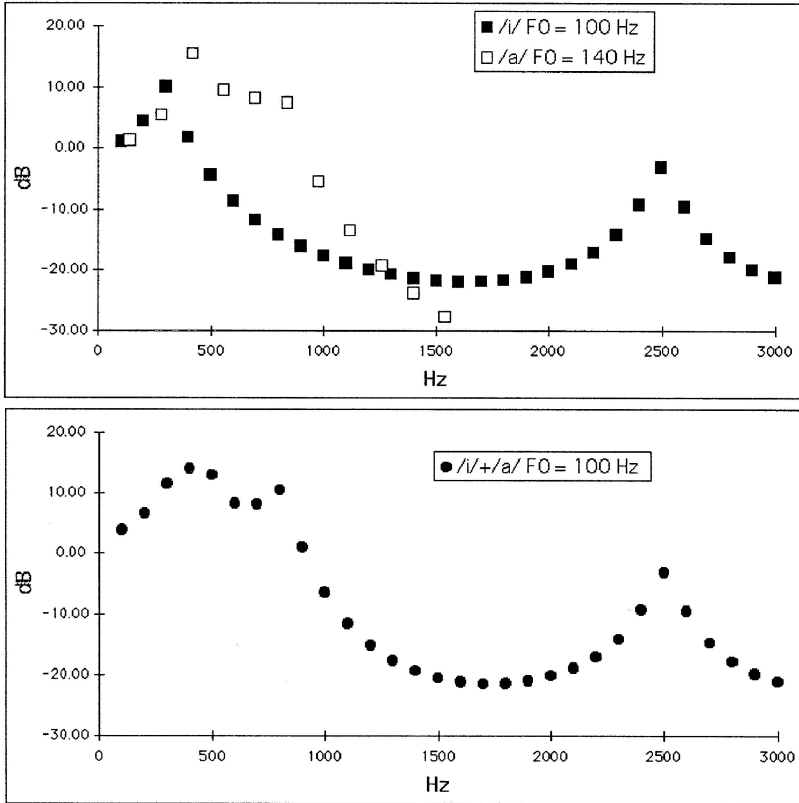


FIGURE 8.6. The *upper panel* shows the individual harmonics of two synthetic vowels:  $/i/$  on an F0 of 100 Hz, with formant frequencies at 300 Hz and 2500 Hz, and  $/a/$  on an F0 of 140 Hz, with formant frequencies at 440 Hz and 800 Hz. The *lower panel* shows the spectrum of the mixture.

different F0 from the other formants. This phonetic segregation is much easier to achieve when the second formant contains resolved harmonics than when it contains only unresolved (Darwin 1992), perhaps reflecting the greater salience of pitch from resolved than from unresolved harmonics (Houtsma and Smurzynski 1990) and the added difficulty of comparing the pitches of resolved and unresolved harmonics (Carlyon and Shackleton 1994; see Plack and Oxenham, Chapter 2).

A second way in which a difference in F0 could help to improve the intelligibility of two simultaneous vowel sounds operates more locally in frequency. When two vowels are on the same F0, each harmonic of their mixture has an amplitude that is simply the vector sum of the two corresponding harmonics from each constituent vowel. The amplitudes of the harmonics of such a mixture are shown in the bottom panel of Figure 8.6. Notice that the two first formants

have now merged into a single broad peak in the spectral envelope. A difference in F0 can thus help to keep separate the formant peaks from the original sounds.

Experiments that clarified which of these two types of process was responsible for the improvement in identification of vowel pairs on different fundamentals were carried out by Culling and Darwin (1993). They constructed chimeric vowels in which the first formant region had a harmonic structure appropriate to one F0, and the higher formants had a harmonic structure appropriate to a different F0. When complementary pairs of such vowels are added together, grouping across formants by a common F0 would result in the inappropriate pairing of the first formant from one vowel with the higher formants from the other vowel. However, within each formant region, there is still a difference in F0 between the two vowels, just as in normally paired vowels that differ in F0. Surprisingly, Culling and Darwin found that their chimeric vowels gave the same sharp improvement in identification as normal vowels when the F0 difference increased from zero to one semitone. Identification of the chimeric vowels deteriorated relative to the normal vowels only when the F0 difference was larger than four semitones. They also found that this pattern of identification persisted even when the difference in F0 between the two vowels was confined to the first-formant region. These results show that for small F0 differences, the improvement in the identification of double vowels is the result of a local F0 difference between the two vowels in the first formant region. It is irrelevant whether there is also an F0 difference in the higher frequencies, or indeed whether a vowel has a consistent F0 throughout its spectrum. However, for large F0 differences, it is important that the low-frequency and high-frequency regions of a vowel have the same F0. The across-formant grouping by F0 envisaged by Broadbent and Ladefoged thus becomes important only for large F0 differences. The asymptotic improvement at one semitone that is seen with normal double vowels is entirely attributable to the local F0 difference within the first formant region.

### 3.1.2 Localization

A difference in the F0 of simultaneous sounds can also help with their localization. We have already seen in Section 2.5 that localization cues can be ineffective for grouping simultaneous sounds. In particular, an ITD gives virtually no improvement in the identification of two simultaneous, steady vowels on the same F0 (Shackleton et al. 1994) or in the identification of the leftmost of two noise-excited vowel-like sounds (Culling and Summerfield 1995). However, if voiced vowels are given a difference in F0 (which itself helps their identification), then an additional difference in ITD of 400  $\mu$ s further improves identification (Shackleton et al. 1994), presumably by giving an additional spatial separation to the two sounds.

More direct evidence that the grouping of sounds by their harmonic relationships is important in localizing complex sounds comes from experiments that have exploited an intriguing effect first noted by Jeffress (1972) and subsequently investigated by Stern et al. (1988). It is well known that a narrow band

of noise centered on 500 Hz ( $f_c$ ) and given an ITD of +1.5 ms ( $t_i$ ) will be heard on the lagging (not the leading) side. Because of phase ambiguity, this stimulus is barely discernible from one that has the complementary ITD of  $-0.5$  ms (i.e.,  $1/f_c - t_i$ ), and the auditory system prefers the shorter ITD. However, Jeffress discovered that if the bandwidth of the sound is gradually increased, while the ITD is maintained at 1.5 ms, then the location of the noise moves across from the lagging to the leading side. Stern et al. replicated this effect and offered an interpretation in terms of the consistency of ITDs across frequency. As additional frequencies are added to the noise the imposed ITD (+1.5 ms) stays constant, but the complementary ITD ( $1/f_c - t_i$ ), being a function of the frequency concerned, varies. The only consistent ITD is then +1.5 ms, and this consistency eventually overcomes the auditory system's preference for short over long ITDs. This phenomenon is interesting since it indicates that ITD information is being integrated across different frequencies in the calculation of lateral position (see also Shackleton et al. 1992). However, it makes sense to perform this integration only across those frequencies that make up a single auditory object—otherwise sounds with different locations could be treated together to give a single average location, rather than separate locations for different objects. Hill and Darwin (1993) showed that harmonicity contributes to this grouping of sounds for across-frequency integration of ITD. They first replicated the Jeffress effect with harmonic sounds—starting with a single frequency component at 500 Hz and then adding additional harmonics of 100 Hz either side of it. As with Jeffress's noise, with the additional harmonics the location changed away from the lagging side toward the leading side. What Hill and Darwin were then able to show was that mistuning the original 500-Hz harmonic by about 3% was sufficient to move it, as a separate sound source back toward the lagging side. In other words, its location was being determined independently of the other frequency components by virtue of its mistuning.

### 3.1.3 Grouping by Frequency Modulation?

Many natural periodic sources change their pitch over a short time scale: the F0 of speech changes continually, while in music individual sung or played notes may have vibrato (at around 6-Hz frequency-modulation rate), and some amount of unpredictable jitter. This frequency modulation (FM) of F0 leads to correlated movement of the constituent harmonics of the sound. Does this movement contribute to auditory grouping?

Somewhat surprisingly, the consensus, at least for vibrato-like movements, is that while a common FM can help to increase the prominence and coherence of a single sound, a difference in FM between two sounds does not help to segregate them.

Various studies speak to the ability of a common pattern of FM to fuse sounds together into a whole. Chowning (1980) found that adding jitter to a synthetic signing voice caused the individual harmonics to fuse more and the voice to sound more natural. In a similar vein, McAdams (1984) played listeners a triplet of synthetic vowels /a/, /i/, /o/ each on a different F0 (adjacent pitches separated

by a perfect fourth), and with various types of vibrato. He asked his listeners to rate the prominence of each vowel and found that giving a target vowel vibrato increased its prominence. Darwin and colleagues (Darwin et al. 1994) examined how the pitch of a complex tone varied with the mistuning of a single harmonic using methods described in Section 2.1. When the whole complex (including the mistuned harmonic) was given a vibrato-like common FM, the mistuned harmonic continued to contribute to the pitch of the complex for larger amounts of mistuning than it did when there was no FM. These studies show that common FM can help to bind together components into a perceptual whole which is more prominent than sounds with a flat pitch contour.

However, a difference in FM does not contribute to the segregation of sound sources independently of any instantaneous difference in F0. In McAdams's experiments, the increase in prominence of a vowel with FM occurred irrespective of whether the other vowels had no vibrato or vibrato that was either correlated or uncorrelated with the target vowel. Uncorrelated vibrato thus did not provide any additional separation of the sounds to that already provided by their substantial static difference in pitch. A similar conclusion was reached by Summerfield and Culling (1992). They synthesized vowels with inharmonic frequency components, so that harmonicity could not group together the components of a vowel, and then imposed coherent FM on these components. Their listeners were unable to use a different pattern of FM to separate a target vowel from a simultaneous masking vowel.

Why is a difference in FM of the F0 of sounds not used to segregate them? Two types of answer have been proposed. First, Carlyon (Carlyon 1991; 1994) has shown that, surprisingly, listeners are unable to tell whether different spectral regions simultaneously contain coherent or incoherent vibrato-like modulation (provided that this distinction is not confounded by changes in harmonicity). In other words, if a group of components in one frequency region is given one type of FM, listeners cannot tell whether the FM applied to another group of components in a different frequency region is coherent with the first FM or phase shifted. This inability may well reflect a lack of specificity in the way the auditory system codes FM phase (Carlyon et al. 2002); the auditory system appears to have a basic limitation in its ability to code the details of frequency modulation. Why might it have failed to evolve such an ability? One possible answer (Carlyon 1992) is that harmonicity together with a general sensitivity to movement provides a strong enough constraint for auditory grouping. Moving harmonics are unlikely to be harmonically related if they are from different sound sources.

## *3.2 Grouping Sounds Sequentially by F0*

### *3.2.1 Streaming of Tones*

The most studied type of auditory streaming is the segregation of a sequence of single sounds into more than one concurrent perceptual stream. The phenom-



enon has been exploited by composers for centuries and is termed “implied polyphony.” Examples occur in Telemann and J.S. Bach’s works for solo recorder or violin. The effect is most simply demonstrated when a high and a low pure tone alternate. When the rate of alternation and the frequency difference between the tones are large enough the single sequence perceptually splits into two streams. A consequence of the splitting is that listeners find it difficult to judge temporal relationships between the two streams, although those within a stream are easy. As Huron (2001) points out, experimental psychologists have periodically rediscovered these effects (Miller and Heise 1950; Heise and Miller 1951; Bozzi and Vicario 1960; Vicario 1960; Schouten 1962; Dowling 1967; Norman 1967; Bregman and Campbell 1971). The extensive parametric work of van Noorden (1975; 1977) established (Fig. 8.7) that when the rate of pre-

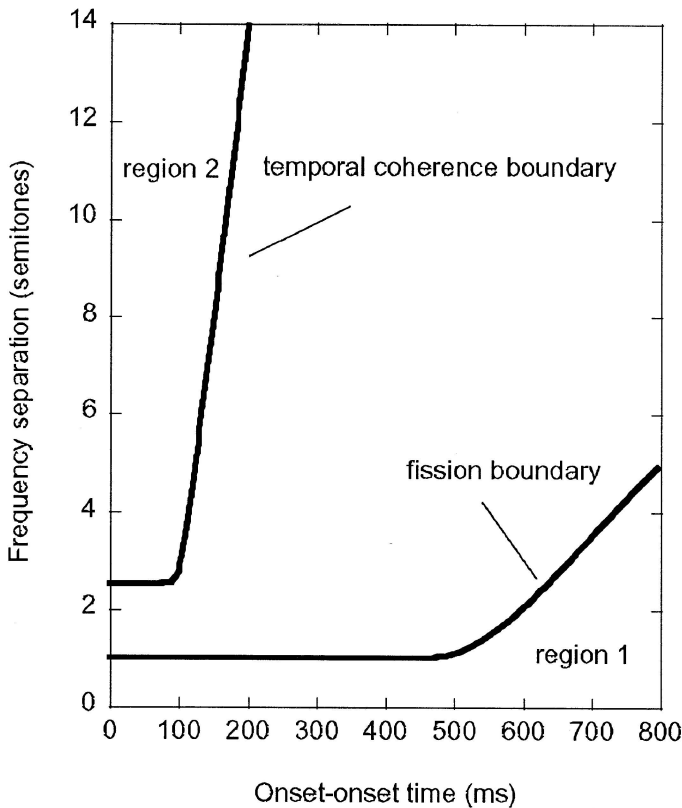


FIGURE 8.7. Boundaries between three different types of percept when listeners hear tones alternating between two frequencies. For very rapid rates of alternation, most frequency differences give a percept of two separate streams (region 2). For very small frequency differences, most rates of alternation give a percept of a single stream (region 1). Between the two the percept is labile and can shift between one or two streams according to a variety of other factors. From Huron (2001), after van Noorden (1977).

sensation is sufficiently slow or the tones sufficiently close in frequency, an alternating sequence of low and high tones is always heard as a single stream (the pitch difference must be a semitone or less for presentation rates faster than two tones/s). If the pitch difference between the tones is greater than about three semitones and the rate of presentation is sufficiently fast (5 to 10 /s depending on the pitch separation) then the sequence is always heard as two separate streams. In between these two extremes lies a region where the percept is labile, and can be influenced by factors such as context. Although van Noorden described his results in terms of the overall rate at which tones were presented (or equivalently the time between the onsets of adjacent tones), it now appears more likely that the important temporal variable is the time between the offset of one tone and the onset of the next tone of similar frequency, that is, the within-stream interstimulus interval (Bregman et al. 2000).

Pure tones of different frequencies differ both in their spectral composition and their pitch. Is this streaming effect due to the pitch of the sound or to its spectral composition? The simplest assumption is that the effect is determined at a relatively peripheral level by the system being loath to alternate rapidly between auditory frequency channels that are distant in frequency (Hartmann and Johnson 1991; Beauvois 1998). However, this type of model based solely on spectral composition cannot explain why segregation occurs for sounds that differ in pitch, but that excite identical auditory filters by virtue of their being composed entirely of unresolved harmonics within the same frequency band (Vliegen and Oxenham 1999). Such streaming occurs even when the listener has to perform a temporal order task which is actually easier when streaming has *not* occurred (Vliegen et al. 1999). A difference in pitch is thus a sufficient condition for auditory streaming, even in the absence of any difference in the sound's auditory spectrum. But it is not a necessary condition. Notes played by orchestral instruments, that have been equated for pitch and loudness, or have insufficient pitch differences to produce streaming, will still stream on the basis of gross spectral differences in timbre and in differences in the duration of a note's attack (Wessel 1979; Iverson 1995).

A difference in the localization of the individual tones in a sequence can also lead to streaming. This streaming can in fact override the streaming that occurs in the mixture through pitch differences. A musical example is given (played on the East African amadinda—a type of xylophone) in Bregman and Ahad's CD of demonstrations of auditory scene analysis (Bregman and Ahad 1995). When two interleaved melodies (alternating notes from each amadinda) are played from the same location, the note mixture streams according to the pitch relationships within the mixture, giving irregular rhythms. However, if the two instruments are sufficiently spatially separated, the previous perceptual organization disappears and each instrument's contribution is heard as a separate stream with a regular rhythm.

In summary, although many perceptual dimensions can lead to sequential streaming (Moore and Gockel 2002), the pitch of a sound is clearly one of them.

### 3.2.2 Attention and Sequential Streaming

An important property of primitive auditory grouping mechanisms as presented by Bregman (Bregman and Rudnick 1975; Bregman 1990) is that they are preattentive—they yield auditory objects that can be the focus of attention. Surprisingly, this hypothesized property was not seriously tested until an important paper by Carlyon and his colleagues (Carlyon et al. 2001). Sequential streaming of alternating high- and low-frequency tones takes a few seconds to build up. If such streaming is preattentive, then build-up should occur even when the tones are not being attended. However, the Carlyon et al. paper showed that almost no buildup of segregation occurs until attention is directed to the tones.

This important result raises the question of how much organization if any takes place on unattended auditory input. Carlyon's result implies that rather little does, and yet musical practice and experimental observations imply that we are capable of maintaining more than two simultaneous streams (one attended, one not). Listeners seem to be able to follow up to three simultaneous instrumental voices in polyphonic music without either underestimating the number of voices or making tracking errors (Huron 1989a); a similar limit applies to estimating the number of notes in a chord (Parncutt 1993). It is also interesting to note (see Fig. 8.8) that in a variety of different polyphonic contexts (ranging from nominally one to five parts) J.S. Bach maintains an average of around three to four auditory streams (Huron 1989b, 2001). Experimentally investigating the effect of attention on the perceptual experience of polyphonic music is an interesting theoretical and practical challenge.

### 3.2.3 Sequential Grouping of Speech by Pitch

When two or more speakers are talking at the same time we can usually follow the voice that we want to listen to without too much interference from the other. The difference in pitch between the two voices and the continuity of the pitch of a particular talker across time help us to achieve this difficult but important feat.

The pitch of the human voice normally changes smoothly over time. In song the pitch changes are more abrupt, but large repeated pitch jumps in melodies are rare (except in yodeling where alternations between chest-voice and falsetto pitches can reach 22 tones/s according to *Guinness World Records 2002* (Young 2001)). A useful heuristic for tracking a voice over time would thus be to track a smoothly changing pitch contour. There is evidence that the human listener does do this. If rapid alternations between a high and a low pitch are introduced into the continuously voiced speech of a single talker, listeners hear the single voice split into two separate voices (one on the lower and one on the higher pitch), with a consequent change in the phonetic content of the speech resulting from the perceptual illusion of silence in each of the two voices, while the other is present (Darwin and Bethell-Fox 1977).

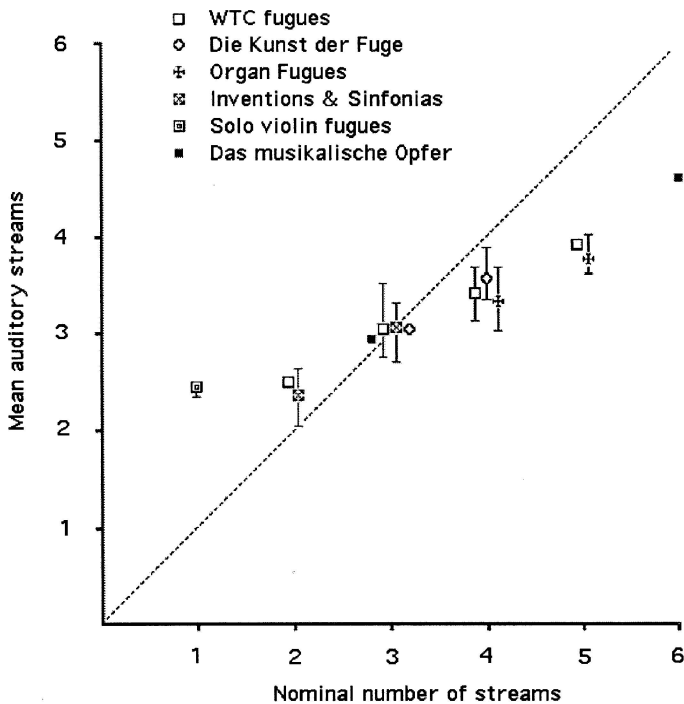


FIGURE 8.8. Mean number of auditory streams according to an algorithm due to Huron (1989b) for a variety of types of polyphonic music by J.S. Bach. Notice that while the nominal number of parts increases, the number of computed streams increases much more slowly and with one exception does not reach four. Reproduced from Huron (2001) with permission of the author and the publishers, University of California Press.

When two voices are naturally present at the same time, a difference in pitch between them will help the listener to disentangle their simultaneous timbres and so to decode the local speech information, as we saw in Section 3.1. But the separate pitch contours also help the listener to track one of the voices over time. This role of pitch has been shown in experiments in which the words that are being spoken are chosen from rather few alternatives, so there is no difficulty for the listener in deciding what individual words have been spoken; poor performance at listening to a particular talker then reflects the listener's inability to follow the voice rather than hear individual words. A difference in pitch between two talkers makes this latter task easier (Darwin and Hukin 2000; Darwin et al. 2003). However, pitch is again not the only cue that can serve this purpose. A difference in location, in overall sound level (Brungart 2001), or in the head sizes of the talkers (Darwin et al., 2003) can also help listeners to track a particular voice.

## 4. Conclusions and Prospect

The relationship between pitch and the perception of mixtures of sound is a rich one, which we are only beginning to understand. In the perception of pitch, some grouping together of the harmonics of a particular sound source by principles of harmonicity and onset time (but not location) seems to be required as a precursor to existing models of pitch perception. A difference in periodicity between two sounds helps listeners to establish which frequency components make up each sound source, and thence to establish their timbres and locations. The generally slowly changing pitch of the voice helps listeners to track the voice of an individual talker over time.

A rich vein for insight into how the brain deals with multiple sound sources lies in principles of musical composition. They codify the accumulated practical experience of composers in achieving the fusion or the segregation of the separate parts in music, and these principles can be related to those that have emerged from the experimental study of sound mixtures (Huron 2001). There is clearly much scope for studies that combine the analytic experimental methods of experimental psychology with composers' practical insight into the behavior of a complex system.

## References

- Assmann PF, Summerfield AQ (1990) Modelling the perception of concurrent vowels: Vowels with different fundamental frequencies. *J Acoust Soc Am* 88:680–697.
- Beauvois MW (1998) The effect of tone duration on auditory stream formation. *Percept Psychophys* 60:852–861.
- Beerends JG, Houtsma AJM (1986) Pitch identification of simultaneous dichotic two-tone complexes. *J Acoust Soc Am* 80:1048–1055.
- Beerends JG, Houtsma AJM (1989) Pitch identification of simultaneous diotic and dichotic two-tone complexes. *J Acoust Soc Am* 85:813–819.
- Bozzi P, Vicario G (1960) Due fattori di unificazione fra note musicali: la vicinanza temporale e la vicinanza tonale. *Rivista di psicologia* 54:253–258.
- Bregman AS (1987) The meaning of duplex perception: sounds as transparent objects. In: Schouten MEH (ed), *The Psychophysics of Speech Perception*. Dordrecht: Martinus Nijhoff, pp. 95–111.
- Bregman AS (1990) *Auditory Scene Analysis: The Perceptual Organisation of Sound*. Cambridge, MA: Bradford Books, MIT Press.
- Bregman AS, Ahad P (1995) *Compact disc: demonstrations of auditory scene analysis*. Montreal: Department of Psychology, McGill University.
- Bregman AS, Campbell J (1971) Primary auditory stream segregation and perception of order in rapid sequences of tones. *J Exp Psychol* 89:244–249.
- Bregman AS, Pinker S (1978) Auditory streaming and the building of timbre. *Canad J Psychol* 32:19–31.
- Bregman AS, Rudnick A (1975) Auditory segregation: stream or streams? *J Exp Psychol Hum Percept Perf* 1:263–267.
- Bregman AS, Ahad PA, Crum PAC, O'Reilly J (2000) Effects of time intervals and tone durations on auditory stream segregation. *Percept Psychophys* 62:626–636.

- Broadbent DE, Ladefoged P (1957) On the fusion of sounds reaching different sense organs. *J Acoust Soc Am* 29:708–710.
- Brokx JPL, Nootboom SG (1982) Intonation and the perceptual separation of simultaneous voices. *J Phon* 10:23–36.
- Brungart DS (2001) Informational and energetic masking effects in the perception of two simultaneous talkers. *J Acoust Soc Am* 109:1101–1109.
- Carlyon RP (1991) Discriminating between coherent and incoherent frequency modulation of complex tones. *J Acoust Soc Am* 89:329–340.
- Carlyon RP (1992) The psychophysics of concurrent sound segregation. *Philos Trans R Soc Lond B* 336:347–355.
- Carlyon RP (1994) Further evidence against an across-frequency mechanism specific to the detection of frequency modulated (FM) incoherence between resolved frequency components. *J Acoust Soc Am* 95:949–961.
- Carlyon RP (1996a) Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker. *J Acoust Soc Am* 99:517–524.
- Carlyon RP (1996b) Masker asynchrony impairs the fundamental-frequency discrimination of unresolved harmonics. *J Acoust Soc Am* 99:525–533.
- Carlyon RP, Shackleton TM (1994) Comparing the fundamental frequencies of resolved and unresolved harmonics: evidence for two pitch mechanisms? *J Acoust Soc Am* 95:3541–3554.
- Carlyon RP, Cusack R, Foxton JM, Robertson RH (2001) Effects of attention and unilateral neglect on auditory stream segregation. *J Exp Psychol Hum Percept Perf* 27:115–127.
- Carlyon RP, Micheyl C, Deeks J, Moore BCJ (2002) A new account of monaural phase sensitivity. *J Acoust Soc Am* 111:2468.
- Chowning JM (1980) Computer synthesis of the singing voice. In Sundberg J (ed), *Sound Generation in Wind, Strings, Computers*. Stockholm: Royal Academy of Music, pp. 4–13.
- Ciocca V, Darwin CJ (1993) Effects of onset asynchrony on pitch perception: adaptation or grouping? *J Acoust Soc Am* 93:2870–2878.
- Culling JF, Darwin CJ (1993) Perceptual separation of simultaneous vowels: within and across-formant grouping by  $F_0$ . *J Acoust Soc Am* 93:3454–3467.
- Culling JF, Summerfield Q (1995) Perceptual separation of concurrent speech sounds: absence of across-frequency grouping by common interaural delay. *J Acoust Soc Am* 98:785–797.
- Darwin CJ (1981) Perceptual grouping of speech components differing in fundamental frequency and onset-time. *Q J Exp Psychol* 33A:185–208.
- Darwin CJ (1992) Listening to two things at once. In Schouten MEH (ed), *The Auditory Processing of Speech: From Sounds to Words*. Berlin: Mouton de Gruyter, pp. 133–147.
- Darwin CJ (1995) Perceiving vowels in the presence of another sound: a quantitative test of the “Old-plus-New” heuristic. In Sorin C, Mariani J, Méloni H, Schoentgen J, (eds), *Levels in Speech Communication: Relations and Interactions: A tribute to Max Wajskop*. Amsterdam: Elsevier, pp. 1–12.
- Darwin CJ, Bethell-Fox CE (1977) Pitch continuity and speech source attribution. *J Exp Psychol Hum Percept Perf* 3:665–672.
- Darwin CJ, Ciocca V (1992) Grouping in pitch perception: effects of onset asynchrony and ear of presentation of a mistuned component. *J Acoust Soc Am* 91:3381–3390.

- Darwin CJ, Carlyon RP (1995) Auditory grouping. In Moore BCJ (ed), *The Handbook of Perception and Cognition*. 2nd ed. Vol. 6: Hearing. London: Academic Press, pp. 387–424.
- Darwin CJ, Hukin RW (1999) Auditory objects of attention: the role of interaural time-differences. *J Exp Psychol Hum Percept Perf* 25:617–629.
- Darwin CJ, Hukin RW (2000) Effectiveness of spatial cues, prosody and talker characteristics in selective attention. *J Acoust Soc Am* 107:970–977.
- Darwin CJ, Ciocca V, Sandell GR (1994) Effects of frequency and amplitude modulation on the pitch of a complex tone with a mistuned harmonic. *J Acoust Soc Am* 95:2631–2636.
- Darwin CJ, Hukin RW, Al-Khatib BY (1995) Grouping in pitch perception: evidence for sequential constraints. *J Acoust Soc Am* 98:880–885.
- Darwin CJ, Brungart DS, Simpson BD Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *J Acoust Soc Am* (2003) 114:2913–2922.
- de Cheveigné A, McAdams S, Laroche J, Rosenberg M (1995) Identification of concurrent harmonic and inharmonic vowels—a test of the theory of harmonic cancellation and enhancement. *J Acoust Soc Am* 97:3736–3748.
- Denbigh PN, Zhao J (1992) Pitch extraction and separation of overlapping speech. *Speech Commun* 11:119–126.
- Dowling WJ (1967) Rhythmic fission and the perceptual organisation of tone sequences. Unpublished doctoral dissertation, thesis. Harvard University, Cambridge, MA.
- Duifhuis H, Willems LF, Sluyter RJ (1982) Measurement of pitch in speech: an implementation of Goldstein's theory of pitch perception. *J Acoust Soc Am* 71:1568–1580.
- Goldstein JL (1973) An optimum processor theory for the central formation of the pitch of complex tones. *J Acoust Soc Am* 54:1496–1516.
- Hartmann WM, Johnson D (1991) Stream segregation and peripheral channeling. *Music Percept* 9:155–183.
- Heise GA, Miller GA (1951) An experimental study of auditory patterns. *Am J Psychol* 64:68–77.
- Hill NI, Darwin CJ (1993) Effects of onset asynchrony and of mistuning on the lateralization of a pure tone embedded in a harmonic complex. *J Acoust Soc Am* 93:2307–2308.
- Houtsma AJM (1984) Pitch salience of various complex sounds. *Music Percept* 1:296–307.
- Houtsma AJM, Smurzynski J (1990) Pitch identification and discrimination for complex tones with many harmonics. *J Acoust Soc Am* 87:304–310.
- Hukin RW, Darwin CJ (1995) Comparison of the effect of onset asynchrony on auditory grouping in pitch matching and vowel identification. *Percept Psychophys* 57:191–196.
- Huron D (1989a) Voice denumerability in polyphonic music of homogeneous timbres. *Music Percept* 6:361–382.
- Huron D (1989b) Voice segregation in selected polyphonic keyboard works by Johann Sebastian Bach. Ph.D. thesis. University of Nottingham, England.
- Huron D (2001) Tone and voice: a derivation of the rules of voice-leading from perceptual principles. *Music Perception* 19:1–64. The Regents of the University of California.
- Iverson P (1995) Auditory stream segregation by musical timbre—effects of static and dynamic acoustic attributes. *J Exp Psychol Hum Percept & Perf* 21:751–763.

- Jeffress LA (1972) Binaural signal detection: vector theory. In Tobias JV (ed), *Foundations of Modern Auditory Theory*, Vol. II. New York: Academic Press, pp. 349–368.
- Kaernbach C, Demany L (1998) Psychophysical evidence against the autocorrelation theory of auditory temporal processing. *J Acoust Soc Am* 104:2298–2306.
- Lieberman AM, Isenberg D, Rakerd B (1981) Duplex perception of cues for stop consonants. *Percept Psychophys* 30:133–143.
- McAdams S (1984) Spectral fusion, spectral parsing and the formation of auditory images. Ph.D. thesis. Stanford University.
- McAdams S, Botte MC, Drake C (1998) Auditory continuity and loudness computation. *J Acoust Soc Am* 103:1580–1591.
- Meddis R, O'Mard L (1997) A unitary model of pitch perception. *J Acoust Soc Am* 102:1811–1820.
- Miller GA, Heise GA (1950) The trill threshold. *J Acoust Soc Am* 22:637–638.
- Moore BCJ (1987) The perception of inharmonic complex tones. In Yost WA, Watson CS (eds), *Auditory Processing of Complex Sounds*. Hillsdale, NJ: Erlbaum, pp. 180–189.
- Moore BCJ, Gockel H (2002) Factors influencing sequential stream segregation. *Acustica* 88:320–333
- Moore BCJ, Glasberg BR, Peters RW (1985a) Relative dominance of individual partials in determining the pitch of complex tones. *J Acoust Soc Am* 77:1853–1860.
- Moore BCJ, Peters RW, Glasberg BR (1985b) Thresholds for the detection of inharmonicity in complex tones. *J Acoust Soc Am* 77:1861–1868.
- Norman D (1967) Temporal confusions and limited capacity processors. *Acta Psychol* 27:293–297.
- Parncutt R (1993) Pitch properties of chords of octave-spaced tones. *Contemp Music Rev* 9:35–50.
- Roberts B, Brunstrom JM (1998) Perceptual segregation and pitch shifts of mistuned components in harmonic complexes and in regular inharmonic complexes. *J Acoust Soc Am* 104:2326–2338.
- Roberts B, Brunstrom JM (2001) Perceptual fusion and fragmentation of complex tones made inharmonic by applying different degrees of frequency shift and spectral stretch. *J Acoust Soc Am* 110:2479–2490.
- Saldanha EL, Corso JF (1964) Timbre cues and the identification of musical instruments. *J Acoust Soc Am* 36:2021–2026.
- Sandell GJ, Darwin CJ (1996) Recognition of concurrently-sounding instruments with different fundamental frequencies. *J Acoust Soc Am* 100:2683.
- Scheffers MT (1979) The role of pitch in perceptual separation of simultaneous vowels. Institute for Perception Research, Annual Progress Report 14:51–54.
- Scheffers MT (1983) Sifting vowels: auditory pitch analysis and sound segregation. Ph.D. thesis, Gröningen University.
- Schouten JF (1962) On the perception of sound and speech. *Proceedings of the 4th International Congress on Acoustics* 2:201–203, ed. Nielsen AK, Copenhagen.
- Shackleton TM, Meddis R, Hewitt MJ (1992) Across frequency integration in a model of lateralisation. *J Acoust Soc Am* 91:2276–2279.
- Shackleton TM, Meddis R, Hewitt MJ (1994) The role of binaural and fundamental frequency difference cues in the identification of concurrently presented vowels. *Q J Exp Psychol* 47A:545–563.
- Stern RM, Zeiberg AS, Trahiotis C (1988) Lateralization of complex binaural stimuli: a weighted image model. *J Acoust Soc Am* 84:156–165.



- Summerfield Q, Culling J (1992) Auditory segregation of competing voices: absence of effects of FM or AM coherence. *Philos Trans Roy Soc Lond B* 336:357–366.
- van Noorden LPAS (1975) Temporal coherence in the perception of tone sequences. Ph.D. thesis. Eindhoven University of Technology.
- van Noorden LPAS (1977) Minimal differences of level and frequency for perceptual fission of tone sequences ABAB. *J Acoust Soc Am* 61:1041–1045.
- Vicario G (1960) Analisi sperimentale di un caso di dipendenza fenomenica tra eventi sonori. *Riv Psicol* 54:83–106.
- Vliegen J, Oxenham AJ (1999) Sequential stream segregation in the absence of spectral cues. *J Acoust Soc Am* 105:339–346.
- Vliegen J, Moore BC, Oxenham AJ (1999) The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task. *J Acoust Soc Am* 106:938–945.
- Wessel DL (1979) Timbre space as a musical control structure. *Comp Mus J* 3:45–52.
- Wightman FL, Kistler DJ (1992) The dominant role of low-frequency interaural time differences in sound localization. *J Acoust Soc Am* 91:1648–1661.
- Woods WA, Colburn S (1992) Test of a model of auditory object formation using intensity and interaural time difference discriminations. *J Acoust Soc Am* 91:2894–2902.
- Young M (2001) Guinness World Records 2002. Gullane.
- Zwicker UT (1984) Auditory recognition of diotic and dichotic vowel pairs. *Speech Commun* 3:265–277.

# Effect of Context on the Perception of Pitch Structures

EMMANUEL BIGAND AND BARBARA TILLMANN

## 1. Introduction

Our interaction with the natural environment involves two broad categories of processes to which cognitive psychology refers as *sensory-driven processes* (also called *bottom-up processes*) and *knowledge-based processes* (also called *top-down processes*). Sensory-driven processes extract information relative to a given signal by considering exclusively the internal structure of the signal. Based on these processes, an accurate interaction with the environment supposes that external signals contain enough information to form adequate representations of the environment and that this information is neither incomplete nor ambiguous. Several models of perception have attempted to account for human perception by focusing on sensory-driven processes. Some of these models are well known in visual perception (Marr 1982; Biederman 1987), as well as in auditory perception (see de Cheveigné, Chapter 6) and, more specifically, music perception (Leman 1995; Carreras et al. 1999; Leman et al. 2000). For example, Leman's model (2000) describes perceived musical structures by considering uniquely auditory images associated with the musical piece. The model comprises of a simulation of the auditory periphery, including outer and middle ear filtering and cochlea's inner hair cells, followed by a periodicity analysis stage that results in pitch images, and that are stored in short-term memory. These pitch patterns are then fed into a self-organizing map that infers musical structures (i.e., keys).

Sensory-driven models have been largely developed in artificial systems. They capture important aspects of human perception. The major problem encountered by these models is that environmental stimuli generally miss some crucial information required for adapted behavior. Environmental stimuli are usually incomplete, ambiguous, and always changing from one occurrence to the next; in addition, their psychological meaning changes as a function of the overall context in which they occur. For example, a small round orange object would be identified as a tennis ball in a tennis court, but as a fruit in a kitchen, and the other way round as an orange in a tennis court when the tennis player

starts to peel it, or as a tennis ball in a kitchen when a child plays with it. A crucial problem for artificial systems of perception consists in formalizing these effects of context on object processing and identification. A fast and accurate adaptation to the everyday-life environment requires the human brain to analyze signals on the basis of what is known about the regular structures of this environment. The cognitive system needs to be flexible in order to recognize a signal despite several modifications of its physical features (as is the case for spoken word comprehension), to anticipate the incoming of future events, to restore missing information, and so on. From this point of view, human brains differ radically from artificial systems by their considerable power to integrate contextual information in perceptual processing. Most of the involved processes are knowledge driven, which results in a smooth interaction with the environment. A further example that highlights the importance of top-down processes is given by considering what happens when something unexpected suddenly occurs in the environment. In some situations, top-down processes are so strong that the cognitive system fails to accomplish a correct analysis of the situation (“I cannot believe my eyes or my ears”). In some contexts, this failure to interpret unexpected events risks being detrimental and may have dramatic consequences (e.g., in industrial accidents).

No doubt, both bottom-up and top-down processes are indispensable for a complete adaptation to the environment. Sensory-driven processes ensure that the cognitive system is informed about the objective structure of the environmental signals, sometimes in a quite automatic way. Top-down processes, by contrast, contribute to facilitate the processing of signals from very low levels (including signal detection) to more complex ones (such as perceptual expectancies or object identification). It is likely that the contribution of both groups of processes depends on several factors relating to the external situation and to the psychological state of the perceiver. For example, in contrast to a silent perceptual setting with clear signals, a noisy environmental situation would encourage top-down process to intervene in order to compensate for the deterioration of the signals. Projective tests used in clinical psychology (e.g., Rorschach test) may be seen as powerful methods to provoke top-down processes for analyzing ambiguous visual figures with the goal of discovering aspects of the individual’s personality. If the visual figures were clearly representing environmental scenes, top-down processes would be less activated.

Although the contribution of top-down processes has been well documented in several domains, including speech perception and visual perception, much remains to be understood about how exactly these processes work in the auditory domain, specifically in nonverbal audition (see McAdams and Bigand 1993). The relatively small part devoted to top-down processes in text books on human audition is rather surprising since no obvious arguments lead us to believe that human audition is more influenced by sensory-driven processes than by top-down processes. The aim of the present and final chapter of this book is to consider some studies that provide convincing evidence about the role played by top-down processes on the processing of pitch structures in music perception.

We start by considering some basic examples in the visual domain, which differentiate both types of processes (see Section 2). We then consider how similar top-down processes influence the perception as well as the memorization of pitch structures (see Section 3) and govern perceptual expectancies (see Section 4). Most of these examples were taken from the music domain. As will become evident in what follows, it is likely that Western composers have taken advantage of the fundamental characteristic of the human brain to process pitch structures as a function of the current context and have thus developed a complex musical grammar based on a very small set of musical notes. Section 5 summarizes some of the neurophysiological bases of top-down processes in the music domain. The last two sections of the chapter analyze the acquisition of knowledge and top-down processes as well as their simulation by artificial neural nets. In Section 6, we argue that regular pitch structures from environmental sounds are internalized through passive exposure and that the acquired implicit knowledge then governs auditory expectations. The way this implicit learning in the music domain may be formalized by neural net models is considered in Section 7. To close this chapter, we put forward some implications of these studies on context effects for artificial systems of pitch processing and for methods of training hearing-impaired listeners (Section 8).<sup>1</sup>

## 2. Bottom-Up versus Top-Down Processes

A first example illustrating the importance of top-down processes in vision is shown in Figure 9.1 and was given by Fisher (1967). Start looking at the left drawing of the first line while masking the second line of the figure. You will identify the face of a man. If now, you look to the other drawings on the right, your perception remains unchanged and the drawing on the extreme right will be perceived as the face of a man. Present now the second line of drawings to another person and require her or him to identify the first drawing on the right, while masking those of the first line. She or he will identify the body of a woman. This perception will not change for the drawings on the left, including the one of the extreme left. The critical point of this demonstration is that the last drawing on the right of the first line is identical to the last drawing on the left of the second line. Nevertheless, the same drawing has been perceived completely differently as a function of the context in which it has been presented. After a set of drawings representing a face, it is identified as a man's face. After a set of drawings representing the body of a woman, it is identified as a body. Since the sensory information is strictly identical in both situations,

---

<sup>1</sup>Music theoretic concepts and basic aspects of pitch processing in music necessary for the understanding of this chapter are introduced in the following sections. Readers interested in more extensive presentations may consult the excellent chapters in Deutsch (1982, 1999) and Dowling and Harwood (1986).

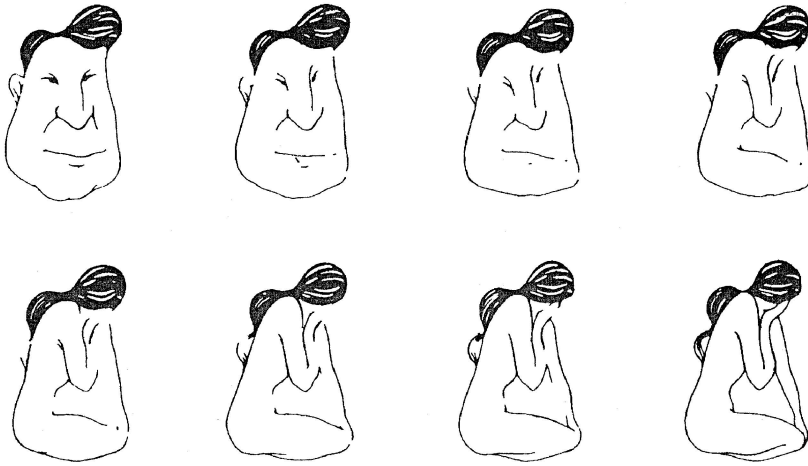
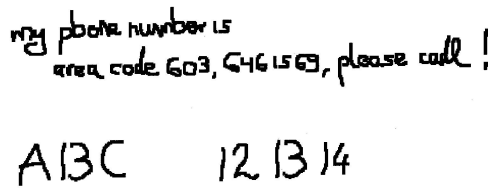


FIGURE 9.1. Example of the importance played by top-down process in vision by Fisher (1967). Reproduced with permission of the Psychonomic Society. See explanations in the text (Section 2).

this difference in perception can be explained by the intervention of top-down, context-dependent, processes that determine perception.

Similar examples are numerous in cognitive psychology, and two further examples are presented here. Just consider the sentence displayed in Figure 9.2 top. If you read, “my phone number is area code 603, 6461569, please call” without any difficulty, some of the letters have been identified differently depending on the word context in which they appear: with the verb “is” being identified as 15 in the code number, the letter “b” as “h” in phone and as “b” in number, and the letter “l” as “d” in code, and as “l” in please. Similar context effects on letter processing have been reported in reading experiments showing that letter identification and memorization is better when letters form meaningful words (word superiority effect). In a related vein, in Figure 9.2 (bottom) you are more likely to interpret the sign in the middle of the two triplets as a B in the sequence on the left and as the number 13 in the sequence on the right. The way a stimulus evolves in space constitutes a further contextual factor that can influence perceptual identification as illustrated by the following example: a hand drawing of a duck can be perceived as representing the flight of a duck when moving from right to left, but as a flight of plane when moving from left to right. Effects of context are not specific to language or vision, and other examples can be found in tasting (Chollet 2001). For example, changing the color of wine is sufficient to identify the wine as red even though it is white wine and vice versa, even among expert wine tasters (Morrot et al. 2001).

Some effects of context have been reported in the auditory domain as well. For example, Ballas and Mullins (1991) reported that the identification of an



my phone number is  
area code 603, 646 15 63, please call !

A13C      12 13 14

FIGURE 9.2. Examples of the importance played by top-down process in reading. See explanations in the text (Section 2). The top figure is adapted from Figure 3.41 in Crider AB, Psychology, 4th ed. © 1993. Reprinted by permission of Pearson Education Inc., Upper Saddle River, NJ.

environmental sound (e.g., a burning detonator) that is acoustically similar to another sound (e.g., food cooking) is weaker when it is presented in a context that biases its identification toward the meaning of the other sound (peeling vegetables/cutting food/a burning detonator) than in a context that is consistent with its meaning (lighting matches/burning detonator/explosion). In a well-known experiment, Warren (1970; Warren and Sherman 1974) reported phonemic restoration effects that depend on the semantic context of the spoken sentence. A phoneme was either removed or replaced by white noise bursts in spoken sentences (indicated by \*). For example: “It was found that the \*eel was on the orange,” “It was found that the \*eel was on the table” or “It was found that the \*eel was on the axle.” As a function of the surrounding sentence, listeners reported hearing “peel,” “meal,” or “wheel” in the three examples. Interestingly, the phenomenon of phonemic restoration only takes place when a noise burst replaces the missing signal. Warren (see Warren 1999 for a review of his work) suggests that a listener hears a sound as being present (participants actually report hearing the phoneme as superimposed on the noise) when there is contextual evidence that the sound may have been present, but has been potentially masked by another sound. Perceptual restoration is not specific to the language domain, and similar effects have been reported in the music domain (Sasaki 1980; DeWitt and Samuels 1990). Sasaki (1980), for example, reported that notes replaced by noise in familiar melodies were “filled in” by the listener. These outcomes suggest that the cognitive system anticipates specific auditory signals on the basis of the previously heard context (either linguistic or musical). This expectancy is strong enough to restore incomplete or missing information. In some cases, the auditory expectations also influence very peripheral auditory processes. Howard et al. (1984) reported that the detection threshold for auditory signals is influenced by a preceding context, even without an explicit signal indicating the pitch height of the to-be-detected target. In their study, a series of sounds constantly decreased in pitch height with the target being the last event. The contextual movement created the expectation that the target would be placed in the continuity, and participants were more sensitive in detecting a target at that expected pitch height.

The influence of context on the processing of pitch structures was reported as early as 1958 by Francès (1958). In one of his experiments, Francès required musicians to detect mistuned notes in piano pieces. This mistuning was performed in different ways. In one condition, some musical notes were mistuned in such a way that the pitch interval between the mistuned notes and those to which they were anchored was reduced.<sup>2</sup> For example, the leading note (the note B in a C major key) is generally anchored to the tonic note (the note C in a C major key). Francès mistuned the note B by increasing its fundamental frequency (F0) so that the pitch interval between the notes B and C (an ascending semitone or half-step) was reduced. In the other experimental condition, this mistuning was performed in the opposite way (the frequency of the B leading tone was decreased). When played without musical context, participants easily perceived both types of mistuning. Placed in a musical context, only the second type of mistuning (which conflicted with musical anchoring) was perceived. This outcome shows that the perceptual ability to perceive changes in pitch structures (in this study, the shift of the F0 of a musical note) is modulated by top-down processes that integrate the function of the note in the overall musical context. It is likely that the effect of top-down processes reported by Francès in this study was driven by listeners' knowledge of Western tonal music. If this experiment was run with listeners who have never been exposed to Western tonal music, these exact context effects may probably not have occurred, or, at least, may have been different (see Castellano et al. 1984).

Since Francès (1958), numerous studies have been performed to further understand the role played by knowledge-driven processes on the perception of pitch structures. Some of these studies demonstrated that the perception and memorization of pitches depend on the musical context in which the pitches appear (see Section 3). During the last decade, several studies provided further evidence that the ease with which we process pitch structures mostly depends on knowledge-driven expectations (see Section 4). We start by reviewing these studies, and we will then consider in more detail whether context effects are hardwired or develop in the brain (see Section 5).

### 3. Effects of Context on the Perception and Memorization of Pitch Structures in Western Tonal Music

Music is a remarkable medium illustrating how top-down and bottom-up processes may be intimately entwined. It is likely that composers initially developed musical syntactic-like rules that took advantage of the psychoacoustic properties of musical sounds. However, these structures have been influenced by centuries of spiritual, ideological, patriotic, social, geographic, and economic

---

<sup>2</sup>In Western tonal music, unstable musical tones instill a tension that is resolved by other specific musical tones in very constrained ways (see Bharucha 1984b, 1996). Unstable tones are said to be anchored to more stable ones.

practices that are not necessarily related to the physical structure of the sound. The music theorist, Rosen (1971), noted that it can be asked whether Western tonal music is a natural or an artificial language. It is obvious that on the one hand, it is based on the physical properties of sound, and on the other hand, it alters and distorts these properties with the sole purpose of creating a language with rich and complex expressive potential. From a historical perspective, the Western harmonic system can be considered as the result of a long theoretical and empirical exploration of the structural potential of sound (Chailley 1951). The challenge for cognitive psychology is to understand how listeners today grasp a system in which a multitude of psychoacoustic constraints and cultural conventions are intertwined. Is the ear strongly influenced by the acoustic foundations of musical grammar, mentally reconstructing the relationship between the initial material and the final system? Or are the combinatorial principals only internal, without a perceived link to the subject matter heard at the time? In the latter case, the perception of pitch (the only musical dimension of interest in this chapter) seems to depend on top-down rather than bottom-up processes. Consider, for example, musical dissonance: Helmholtz (1885/1954) postulated that dissonance is a sensation resulting from the interference of two sound waves close in frequency, which stimulate the same auditory filter in conflicting ways. Although it is linked to a specific psychoacoustic phenomenon, this sensation of dissonance relies on a relative concept that cannot explain the structure of Western music on its own (cf. Parncutt 1989). The idea of dissonance has evolved during the course of musical history: certain musical intervals (e.g., the 3rd) were not initially considered as consonant. Each musical style could use these sensations of dissonance in many ways. For example, a minor chord with a major 7th is considered to be perfectly natural in jazz, but not in classical music. Similarly, certain harmonic dissonances of Beethoven, whose musical significance we now take for granted, were once considered to be harmonic errors that required correction (cf. Berlioz 1872). Even more illustrative examples of the cultural dimension of dissonance are innumerable when considering contemporary music or the different musical systems of the world. These few preliminary notes show that sensory qualities linked to pitch cannot be understood outside of a cultural reference frame.

It is actually well established in the music cognition domain that a given auditory signal (a musical note) can have different perceptual qualities depending on the context in which it appears. This context dependency of musical note perception was exhaustively studied by Krumhansl and collaborators from 1979 to 1990 (for a summary of this research see Krumhansl 1990). To understand the rationale of these studies, let us consider shortly the basic structures of the Western musical system.

Two aspects of the notion of pitch can be distinguished in music: one related to the fundamental frequency  $F_0$  of a sound (measured in Hertz), which is called pitch height, and the other related to its place in a musical scale, which is called pitch chroma. Pitch height varies directly with frequency over the range of the audible frequencies. This aspect of pitch corresponds to the sensation of high



and low. Pitch chroma embodies the perceptual phenomenon of octave equivalence by which two sounds separated by an octave are perceived as somewhat equivalent. Pitch chroma is organized in a circular fashion, with octave-equivalent pitches considered to have the same chroma. Pitches having the same chroma define pitch classes. In Western music, there are 12 pitch classes referred to with the following labels: C, C# or Db, D, D# or Eb, E, F, F# or Gb, G, G# or Ab, A, A# or Bb, and B. All musical styles of Western music (from baroque music to rock 'n roll and jazz music) rest on possible combinations of this finite set of 12 pitch classes. Figure 9.3 illustrates the most critical features of these pitch classes combined in the Western tonal system.

The specific constraints to combine these pitch classes have evolved through centuries and vary as a function of stylistic periods. The basic constraints that are common to most Western musical styles are described in textbooks of Western harmony and counterpoint. A complete description of these constraints is beyond the scope of this chapter, and we will simply focus on those features that are indispensable for understanding the basis of context effects in Western tonal music. For this purpose, it is sufficient to understand that the 12 pitch classes are combined into two categories of musical units: chords and keys. The musical notes (i.e., the 12 chromatic notes) are combined to define musical

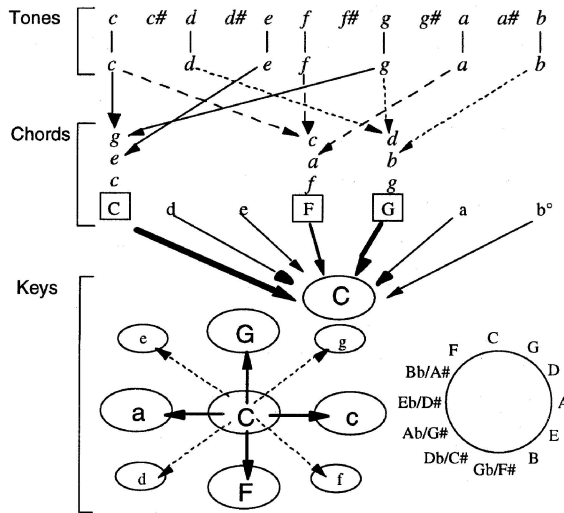


FIGURE 9.3. Schematic representation of the three organizational levels of the tonal system. (*Top*) Twelve pitch classes, followed by the diatonic scale in C major. (*Middle*) Construction of three major chords, followed by the chord set in the key of C major key. (*Bottom*) Relationships of the C major key with close major and minor keys (*left*) and with all major keys forming the circle of fifths (*right*). (Tones are represented in *italics*, minor and major chords/keys in *lower-* and *uppercase*, respectively.) From Tillmann et al. (2001).

chords. For example, the notes C, E and G define a C major chord, and the notes F, A and C define an F major chord. The frequency ratios between two notes define musical pitch intervals and are expressed in the music domain by the number of semitones (for a presentation of intervals in terms of frequency ratios see Burns 1999, Table 1). For example, the distance in pitch between the notes C and E is four semitones and defines the pitch interval of a major 3rd. The pitch interval between the notes C and *E<sub>b</sub>* is three semitones, and defines a minor 3rd. The pitch interval between the notes C and G is seven semitones, and defines a perfect 5th. A diminished 5th is defined by two musical notes separated by six semitones (e.g., C and *G<sub>b</sub>*). Musical chords can be major, minor, or diminished depending on the types of interval they are made of. A major chord is made of a major 3rd and a perfect 5th (e.g., C–E, and C–G, respectively). A minor chord is made of a minor 3rd and a perfect 5th (e.g., C–*E<sub>b</sub>* and C–G). A diminished chord is made of a minor 3rd (C–*E<sub>b</sub>*) and diminished 5th (e.g., C–*G<sub>b</sub>*). A critical feature of Western tonal music is that a musical note (say C) may be part of different chords (e.g., C, F, and *A<sub>b</sub>* major chords, c, a, and f minor chords), and its musical function changes depending on the chord in which it appears. For example, the note C acts as the root, or tonic, of C major and c minor chords, but as the dominant note in F major and f minor chords.

The 12 musical notes are combined to define 24 major and minor chords that, in turn, are organized into larger musical categories called musical keys. A musical key is defined by a set of pitches (notes) within the span of an octave that are arranged with certain pitch intervals among them. For example, all major keys are organized with the following scale: two semitones (C–D in the case of the C major key), two semitones (D–E), one semitone (E–F), two semitones (F–G), two semitones (G–A), two semitones (A–B), and one semitone (B–C'). The scale pattern repeats in each octave. By contrast, the minor keys (in its minor harmonic form) are organized with the following scale: two semitones (C–D, in the case of the C minor key), one semitone (D–*E<sub>b</sub>*), two semitones (*E<sub>b</sub>*–F), two semitones (F–G), one semitone (G–*A<sub>b</sub>*), three semitones (*A<sub>b</sub>*–G), and one semitone (B–C). On the basis of the 12 musical notes and the 24 musical chords, 24 musical keys can be derived (e.g., 12 major and 12 minor keys).<sup>3</sup> For example, the chords C, F, G, d, e, a, and *b<sup>°</sup>* belong to the key of C major, and the chords F#, C#, B, *g<sup>#</sup>*, *a<sup>#</sup>*, *d<sup>#</sup>*, and *e<sup>#°</sup>* define the key of F# major. Further structural organizations exist inside each key (referred to as tonal-harmonic hierarchy in Krumhansl, 1990) and between keys (referred to as interkey distances). The concept of tonal hierarchy designates the fact that some musical notes have more referential functions inside a given key than others.

---

<sup>3</sup>The first attempt to musically explore all of these keys was done by J.S. Bach in the Well-Tempered Clavier. Major, minor, and diminished chords are defined by different combinations of three notes. Minor chords and minor keys are indicated by lowercase letters, and major chords and major keys by uppercase letters. The symbol ° refers to diminished chords.

The referential notes act in the music domain like cognitive reference points act in other human activities (Rosch 1975, 1979). Human beings generally perceive events in relation to other more referential ones. As shown by Rosch and others, we perceive the number 99 as being almost 100 (but not the reverse), and we prefer to say that basketball players fight like lions (but not the reverse). In both examples, “100” and “lion” act as cognitive reference points for mental representations of numbers and fighters (see also Collins and Quillian 1969). Similar phenomena occur in music. In Western *tonal* music, the tonic of the key is the most referential event in relation to which all other events are perceived (Schenker 1935; Lerdahl and Jackendoff 1983, for a formal account).<sup>4</sup> Supplementary reference points exist, as instantiated by the dominant and mediant notes.<sup>5</sup> These differences in functional importance define a within-key hierarchy for notes. A similar hierarchy can be found for chords: chords built on the first degree of the key (the tonic chord) act as the most referential chord of Western harmony, followed by the chords built on the 5th and 4th scale degrees (called dominant and subdominant, respectively).

Intrakey hierarchies are crucial in accounting for context effects in music. Indeed, a note (and also a chord) has different musical functions depending on the key context in which it appears. For example, the note C acts as a cognitive reference note in the C major and c minor keys, as the less referential dominant note in the F major and minor keys, as a moderate referential mediant note in the Ab major key and the a minor key, as weakly referential notes in the major keys of Bb, G, and Eb as well as in the minor keys of bb, g, and e, as an unstable leading note in the major and minor keys of Db and as nonreferential, nondiatonic note in all remaining keys. As the 12 pitch classes have different musical functions depending on the 12 major and 12 minor key contexts in which they can occur, there are numerous possibilities to vary the musical qualities of notes in Western tonal music. The most critical feature of the Western musical system is thus to compensate for the small number of pitch classes (12) by taking advantage of the influence of context on the perception of these notes.

---

<sup>4</sup>The tonal system refers to a set of rules that characterize Western music since the Baroque (17th century), Classical, and Romantic styles. This system is still quite prominent in the large majority of traditional and popular music (rock, jazz) of the Western world as in Latin America.

<sup>5</sup>Western music is based on an alphabet of 12 tones, known as the chromatic scale. This system then constitutes subsets of seven notes from this alphabet, each subset being called a scale or key. The key of C major (with the tones C, D, E, F, G, A, B) is an example of one such subset. The first, third, and fifth notes of the major scale (referred to as tonic, median, and dominant notes) act as cognitive reference notes. Musical chords correspond to the simultaneous sounding of three different notes. A chord is built on the basis of a tone, which is called the root and gives its name to the chord, so that the C major chord corresponds to a major chord built on the tone C. In a given key, the chords built on the first, fourth, and fifth notes of the scale (i.e., C, F, and G, in a C major scale, for example) are referred to as tonic, subdominant, and dominant chords. These chords act as cognitive reference events in Western music (see Krumhansl 1990 and Bigand 1993 for reviews).

In other words, there are 12 physical event classes in Western music, but since these events have different musical functions depending on the context in which they occur, the Western tonal system has a great number of possible musical events.

A further way to understand the importance of this feature for music listening is to consider what would happen if the human brain were not sensitive to contextual information. All the music we listen to would be made of the same 12 pitch classes. As a result, there would be a huge redundancy in pitch structures inside a given musical piece as well as across all Western musical pieces. As a consequence, we may wonder whether someone would enjoy listening to Beethoven's 9th symphony, Dvorak's *Stabat Mater*, or Verdi's *Requiem* until the end of the piece (with a duration of about 90 minutes) and whether someone would continue to enjoy listening to these musical pieces after having perceived them once or twice.<sup>6</sup> This problem would be even more crucial for absolute pitch listeners who are able to perceive the exact pitch value of a note without any reference pitch. It is likely that composers have used the sensitivity of the human brain for context effects in order to reduce this redundancy. Indeed, Western musical pieces rarely remain in the same musical key. Most of the time, several changes in key occur during the piece, the number of changes being related to the duration of the piece. These key changes modify the musical functions of the notes and result in noticeable changes of the perceptual qualities of the musical flow. For a very long time, Western composers have used the psychological impact of these changes in perceptual qualities for expressive purposes (see Rameau 1721 for an elegant description). Expressive effects of key changes or modulations are stronger when the second key is musically distant from the previous one. For example, the changes in perceptual qualities of the musical flow resulting from the modulation from the key of C major to the key of G major will be moderate and less salient than those resulting from a modulation from the C major key to the F# major key.

The musical distances between keys are defined in part by the number of notes (and chords) shared by the keys. For example, there are more notes shared by the keys of C and G major than by the keys of C and F# major. A simplified way to represent the interkey distances is to display keys on a circle (Fig. 9.2, bottom), which is called the circle of fifths. Major keys are placed on this circle as a function of the number of shared notes (and chords), with more notes and chords in common between adjacent keys on the circle. Interkey distances with minor keys are more complex to represent because the 12 minor keys share different numbers of notes and chords with major keys. Moreover, the number of shared notes and chords defines only a very rough way to describe musical distances between keys. A more convincing way to compute these distances

---

<sup>6</sup>To some extent, 12-tone music of Schoenberg, Webern, and Berg faces this difficult problem when using rows of 12 pitch classes for composing long musical pieces without the possibility to manipulate their musical function. Not surprisingly, the first dodeca-phonic pieces were of very short duration (see Webern pieces for orchestra).

considers the strength of the changes in musical functions that occurs for each note and chord when the music modulates from one key to another (see Lerdahl 1988, 2001; Krumhansl 1990). A complete account of this computation is beyond the scope of this chapter, but one example is sufficient to explain the underlying rationale. The number of notes shared by the C major key and the c minor key is five (i.e., the notes C, D, F, G, and B). The number of notes shared by the C major key and the B $\flat$  major key is also five (i.e., C, D, F, G, and A). Nevertheless, the musical distance between the former keys is less strong than between the latter keys. This is because the change in musical functions are less numerous in the former case than in the latter. Indeed, the cognitive reference points (tonic and dominant notes) are the same (C and G) in the C major and c minor key contexts. By contrast, these two notes are not referential in the key context of B $\flat$  major (in which the notes B $\flat$  and F act as the most referential notes). As a consequence, a modulation from the C major key to the B $\flat$  major key has more musical impact than a modulation toward the c minor key. More generally, by choosing to modulate from one key to another, composers modify the musical functions of notes, which results in expressive effects for Western listeners: the more distant the musical keys are, the stronger the effect of the modulation. Composers of the Romantic period (e.g., Chopin) used to modulate more often toward distant keys than did composers of the Baroque (e.g., Vivaldi, Bach) and Classical periods (e.g., Haydn, Mozart). If human brains were not integrating contextual information for the processing of pitch structures, all these refinements in musical styles would probably have never been developed.

To summarize, the most fundamental aspect of Western music cognition is to understand the context dependency of musical notes and chords and of their musical functions. Krumhansl's research provides a deep account of this context dependency of musical notes for both perception and memorization. In her seminal experiment, she presented a short tonal context (e.g., seven notes of a key or a chord) followed by a probe note (defining the "probe-note" method). The probe note was one note of the 12 pitch classes. Participants were required to evaluate on a seven-point scale how well each probe note fit with the previous context. As illustrated in Figure 9.4, the goodness-of-fit judgments reported for the 12 pitch classes varied considerably from one key context to another. Musical notes receiving higher ratings are said to be perceptually stable in the current tonal context. Krumhansl and Kessler's (1982) tonal key-profiles demonstrated that the same note results in different perceptual qualities, referred to as musical stabilities, depending on the key of the tonal context in which it appears. These changes in musical stability of notes as a function of key contexts can be considered as the cognitive foundation of the expressive values of modulation.

Krumhansl also demonstrated that within-key hierarchies influence the perception of the relationships between musical notes. In her experiments, pairs of notes were presented after a short musical context and participants rated on a scale from 1 to 7 the degree of similarity of the second note to the first note,

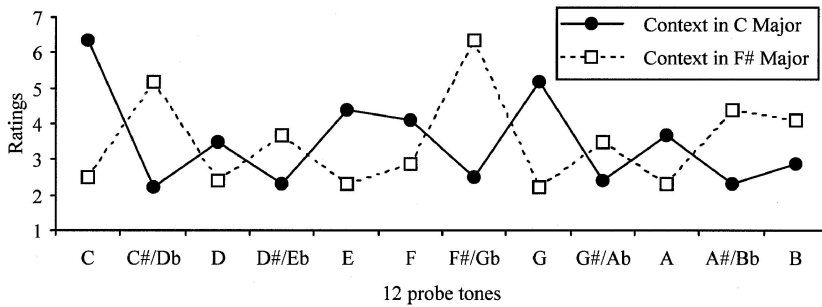


FIGURE 9.4. Probe tone ratings for the 12 pitch classes in C major and F# major contexts. From Krumhansl and Kessler (1982). Adapted with permission of the American Psychological Association.

given the preceding tonal context. All possible note pairs were constructed with the 12 pitch classes. The note pairs were presented after short tonal contexts that covered all 24 major and minor keys. The similarity judgments can be interpreted as an evaluation of the psychological distance between musical notes with more similarly judged notes corresponding to psychologically closer notes. The critical point of Krumhansl's finding was that the psychological distances between notes depended on the musical context as well as on the temporal order of the notes in the pair. For example, the notes G and C were perceived as being closer to each other when they were presented after a context in the C major key than after a context in the A major key or the F# major key. In the C major key context, the G and C notes both act as strong reference points (as dominant and tonic notes, respectively) which is not the case in the A and F# major keys to which these notes do not belong.

This finding suggests that musical notes are perceived as more closely related when they play a structurally significant role in the key context (i.e., when they are tonally more stable). In other words, tonal hierarchy affects psychological distances between musical pitches by a principle of *contextual distance*: the psychological distance between two notes decreases as the stability of the two notes increases in the musical context. The temporal order of presentation of the notes in the pair also affected the psychological distances between notes. In a C major context for example, the psychological distance between the notes C and D was greater when the C note occurred first in the pair than the reverse. This *contextual asymmetry principle* highlights the importance of musical context for perceptual qualities of musical notes and shows the influence of a cognitive representation on the perception of pitch structures.

A further convincing illustration of the influence of the temporal context on the perception of pitch structures was reported by Bharucha (1984a). In one experimental condition, he presented a string of musical notes, such as B3–C4–D#4–E4–F#4–G4, to the participants. In the other experimental condition, the

temporal order of these notes was reversed leading to the sequence G4–F#4–E4–D#4–C4–B3. In the musical domain, this sequence is as ambiguous as the well-known Rubin figure in the visual domain, which can be perceived either as a goblet or two faces. Indeed, the sequence is based on the three notes of the C major chord (C–E–G) that are interleaved with the three notes of the B major chord (B–D#–F#). Interestingly, these chords do not share a parent key, and are thus somewhat incompatible. Bharucha demonstrated that the perception of this pitch sequence depends on the temporal order of the pitches. Played in the former order, the sequence is perceived as being in C major; played in the latter order, it is perceived in B major. In other words, the musical interpretation of an identical set of notes changes with the temporal order of presentation. This effect of context might be compared with the context effect described above concerning the influence of stimulus movement on visual identification (duck versus plane).

The context effects summarized in the preceding discussion have also been reported for the memorization of pitch structures. For example, Krumhansl required participants to compare a standard note played before a musical sequence to a comparison note played after this musical sequence. The performance in this memorization task depended on the musical function of both standard and comparison notes in the interfering musical context. When standard and comparison notes were identical (i.e., requiring a *same response*), performance was best when the notes acted as the tonic note in the interfering musical context (e.g., C in the C major key), it diminished when the notes acted as mediant (e.g., E in the C major key) and was worst when they did not belong to the key context. This finding underlines the role of the *contextual identity principle*: The perception of identity between two instances of the same musical note increases with the musical stability of the note in the tonal context. When standard and comparison notes were different (i.e., requiring a *different response*), the memory errors (confusions) also depended on the musical function of these notes in the interfering musical context, as well as on the temporal order. For example, when the comparison note acted as a strong reference note in the context (e.g., a tonic note) and the standard as a less referential note, memory errors were more numerous than when the comparison note acted as a less referential note and the standard as a strong reference note in the context. This finding cannot be explained by sensory-driven processes. It suggests that in the auditory domain, as in other domains (see, e.g., Rosch for the visual domain), some pitches act as cognitive reference points in relation to which other pitches are perceived. It thus provides a further illustration of the principle of contextual asymmetry described above. Consistent support for contextual asymmetry effects on memory was reported by Bharucha (1984a,b) with a different experimental setting.

Several attempts have been made to challenge Krumhansl and colleagues' demonstration of the cognitive foundation of musical pitch. For example, Huron and Parncutt (1993) argued that most of Krumhansl's probe-note data may be accounted for by a sensory model and can emerge from an echoic memory

model based on pitch salience and including a temporal decay parameter. More recently, Leman (2000) provided a further challenge to these data arguing that none of the previously reported context effects occur at a cognitive level but may simply be explained by some sort of sensory priming. Notably, Leman (2000) simulated data with the help of a short-term memory model based on echoic images of periodicity pitch only.

Given that both top-down and bottom-up processes are intimately entwined in Western music, a critical issue remains to assess the strength of each type of process for music perception. Dowling's remarkable work has demonstrated how both processes may contribute to melodic perception and memorization (Dowling 1972, 1978, 1986, 1991; Bartlett and Dowling, 1980, 1988; Dowling and Bartlett, 1981; Dowling et al. 1995). The influence of bottom-up processes is reflected by listeners' sensitivity to the melodic contour (that is the up-and-down of pitch intervals in the melody). Top-down influences are reflected by the importance of the position of the notes in the musical scale (e.g., tonic or dominant). One critical feature of Dowling's experiments was to demonstrate that a change in melodic contour was more difficult to perceive when the comparison melody was played in a far rather than a close key. A further fascinating finding of Dowling was to show that a given melody played in two different harmonic contexts was not easily perceived as having exactly the same melodic contour. The change in scalar position of the melodic notes from one musical key context to the other interfered with the ability to perceive the melodic contour.

One of our experiments on melody perception directly addressed the strength of top-down processes in a very similar way (Bigand 1997). The study involved presenting 29-note sequences (Figure 9.5) to participants. The challenge was to modify the perception of these note sequences by changing only a few pitches (i.e., five pitches between melody T1 and melody T2). On music theoretical grounds, these few pitch changes should be sufficient to make participants perceive the melody T1 in the context of an a minor key and the melody T2 in the context of a G major key. Given that the musical stability of individual notes changes as a function of key, the profile of perceived musical stability was supposed to vary strongly from T1 to T2, even though both melodies shared a large set of pitches, the same contour and the same rhythm. For example, stop note 2 is a strong referential tonic note in T1, but a weak referential subtonic note in T2. Similarly, stop note 4 is a rather referential mediant note in T1 and a less referential subdominant note in T2. By contrast, stop note 3 is a weak referential supertonic in T1, but a rather strong referential mediant in T2. Readers familiar with music can observe that notes that are referential in one melodic context are less referential in the other, and this is valid up to the last note. Indeed, stop note 23 is a referential tonic in T1, but a less referential supertonic in T2. As a consequence, melody T1 sounds complete, but melody T2 does not. The experimental method to measure perceived musical stability consisted in breaking the melody into 23 fragments, each starting from the beginning of the melody and ending on a different note of the melody (i.e., incremental



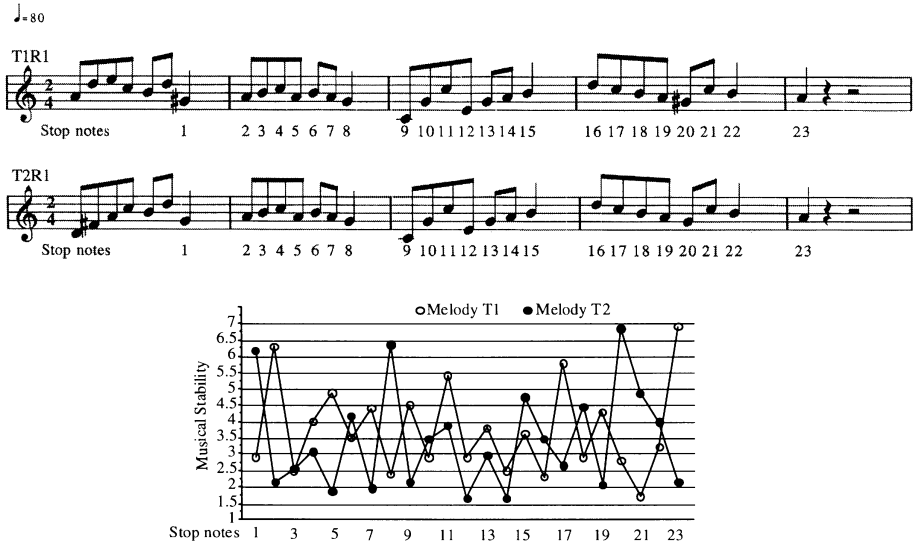


FIGURE 9.5. (*Top*) The two melodies T1 and T2 used in Bigand (1996) with their 23 stop notes on which musical stability ratings were given by participants. (*Bottom*) Musical stability ratings from musician participants superimposed on the two melodies T1 and T2. From Bigand (1996), Fig. 2. Adapted with permission of the American Psychological Association.)

method). As in Krumhansl and Palmer (1987a,b) studies, participants were required to evaluate the degree of completeness of each fragment. Fragments ending on a stable musical note were supposed to result in stronger feelings of musical completion than those ending on a musically instable note. As a consequence, we predicted musical stability profiles to vary strongly from T1 to T2.

The observed stability profiles of the two melodies were negatively correlated in both musicians' and nonmusicians' data (see Fig. 9.5, bottom, for musicians' data). This outcome shows that listeners (musician and nonmusicians) perceived the pitch structure of the two melodies differently, even though they largely contained the same set of pitches and pitch intervals, and had identical melodic contours and rhythms. Moreover, when these melodies were used in a memorization task, participants estimated on average that about 50% of the pitches of the T2 melodies had been changed to create the T1 melodies (Bigand and Pineau 1996). Surprisingly, musicians did not outperform nonmusicians in this task suggesting that for both groups of listeners the musical functions of melodic notes contributed more strongly to defining the perceptual identity of a melody than the actual pitches, pitch intervals, melodic contour and rhythm. Both studies underline the strength of cognitive top-down processes on the perception and memorization of melodic notes.

As explained above, musical notes define the smallest building block of Western tonal music. Musical chords define a larger unit of Western musical pitch structures. A musical chord is defined by the simultaneous sounding of at least three notes, one of these notes defining the root of the chord. Other notes may be added to this triadic chord, which results in a large variety of musical chords. The influence of musical context on the perception of the musical qualities of these chords, as well as the perceptual relationships between these chords has been largely investigated by Krumhansl and collaborators (see Krumhansl 1990 for a summary). The rationale of these studies follows the rationale of the studies briefly summarized above for musical notes (see Krumhansl 1990).

For example, in Bharucha and Krumhansl (1983), two chords were played after a musical context, and participants rated on a seven-point scale the similarity of the second chord to the first one given the preceding context. The pairs of chords were made of all combinations of chords belonging to two musical keys that share only a few pitches (C and F# major). In other words, these keys are musically very distant. If the perception of harmonic relationships was not context dependent, the responses of participants would not have been affected by the context in which these pairs were presented. Figure 9.6 demonstrates that the previous musical context had a huge effect on the perceived relationships of the two chords. When the context was in the key of C major, the chords of the C major key were perceived as more closely related than those of the F# major key. When the F# major key defined the context, the inverse phenomenon was reported. The most critical finding was that when the musical key of the context progressively moved from the C major key to the F# major key through the keys of G, A, and B (see the positions of these keys on the cycle of fifths, Fig. 9.3), the perceptual proximity between the chord pairs progressively changed, so that C major chords progressively were perceived as less related, and F# major chords more related (cf. Krumhansl et al. 1982b). Similar context effects have also been reported in memory experiments, suggesting that it is unlikely that these context effects are caused by sensory-driven processes solely (Krumhansl 1990; Bharucha and Krumhansl 1983).

It is difficult to rule out entirely the influence of sensory-driven processes on the perception of Western harmony in these experiments. This restriction applies even though the authors carefully used Shepard tones (Shepard 1964)<sup>7</sup> and provided converging evidence from perceptual and memory tasks, which suggests that the reported context effects occurred at a cognitive level. The purpose of one of our studies was to contrast sensory and cognitive accounts of the perception of Western harmony (Bigand et al. 1996). Participants listened to triplets of chords with the first and third chords being identical (e.g., X–C–X). Only

---

<sup>7</sup>Shepard tones consist, for example, of five sine wave components spaced at octave frequencies in a five-octave range with an amplitude envelope being imposed over this frequency range so that the components at low and high ends approach hearing threshold. These tones have an organ-like timbral quality and minimize the perceived effect of pitch height.

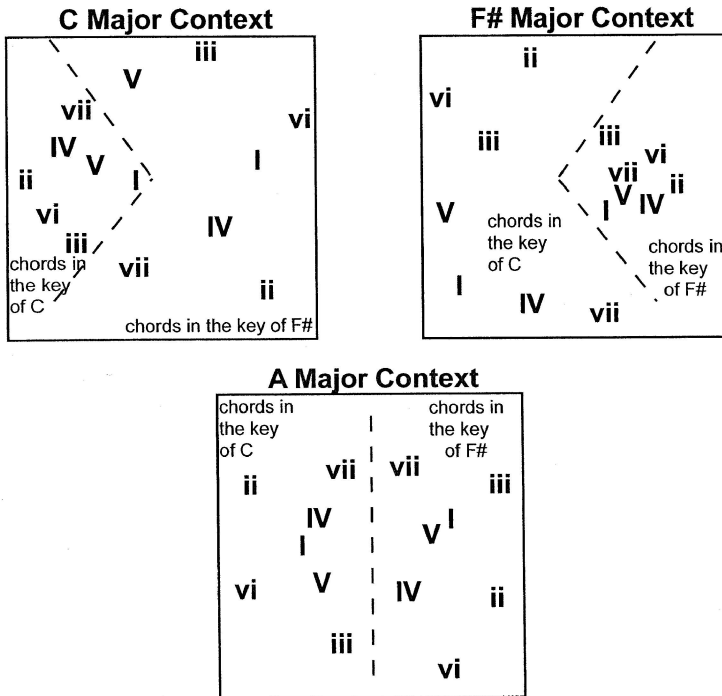


FIGURE 9.6. Representations based on chord similarity ratings in the contexts of C major, F# major and A major. Reprinted from *Cognition*, 13, Bharucha and Krumhansl, The representation of harmonic structure in music: hierarchies of stability as a function of context, pp. 63–102. Copyright (1983), with permission from Elsevier; and from *Perception & Psychophysics*, 32, Krumhansl et al. Key distance effects on perceived harmonic structure in music, 96–108 Copyright (1982) with permission from Psychonomic Society. The closer chords are in the plane, the more similar they are rated to be. Roman numbers refer to the functions of the chords in the key. They reflect the degree of the scale on which the chords are constructed, for example, I for tonic, IV for subdominant, V for dominant, and ii, iii, vi, and vii for chords constructed on 2nd, 3rd, 6th, and 7th degrees of the scale.

the second chord was manipulated and participants evaluated on a 10-point scale the musical tension instilled by the second chord. The manipulated chord was either a triad (i.e., the 12 major and 12 minor triads) or a triad with a minor seventh (i.e., 12 major chords with minor seventh, and 12 minor chords with a minor seventh). The musical tensions were predicted by Lerdahl's cognitive tonal pitch space theory (Lerdahl 1988) and by several psychoacoustical models, including Parncutt's theory (Parncutt 1988). One of the main outcomes was that all models contributed to predicting the perceived musical tension, with albeit a stronger contribution of the cognitive model. This outcome suggests that the abstract knowledge of Western pitch regularities constitutes some kind of cog-

nitive filter that influences how we perceive musical notes and chords. A further influence of this knowledge is documented in the next section by showing that internalized pitch regularities also result in the formation of perceptual expectancies that can facilitate (or not) the processing of pitch structures.

#### 4. Influence of Knowledge-Driven Expectancy on the Processing of Pitch Structures

Once we are familiarized with a given environment, we process environmental stimuli in a highly constrained way. For example, we are not able to ignore linguistic information displayed in our native language, and we automatically anticipate from a previous context the type of events that are likely to occur next. Irrepressible processing and perceptual anticipation have been documented in a variety of domains, including language, face processing and vision. During the last decade, numerous studies have been devoted to investigating the influence of auditory expectations on the processing of pitch structures in the music domain. The seminal studies on harmonic expectancies involved very short contexts. For example, in Bharucha and Stoeckig (1987), participants were required to perform a simple perceptual task on a target chord that was preceded by a prime chord. The harmonic relationship between the prime chord and the target chord defined the variable of interest, and the critical point was to assess whether this relationship influenced the processing of the target. For the purpose of the experimental task, the target chord was either in tune or out of tune, and participants had to decide quickly and accurately whether the target was in tune or out of tune. The principal outcome was that the processing of in-tune targets (e.g., a C major chord) was easier and faster when the target was preceded by a musically related prime chord (e.g., a G major chord) than by a musically unrelated prime chord (e.g., an F# major chord). In the research of Bharucha and collaborators, the effect of context was reversed for out-of-tune targets (with better identification of out-of-tune targets when preceded by a musically unrelated prime). These findings provided evidence for the anticipatory processes that occur from chord to chord when listening to music.

Further experiments were performed to confirm that priming effects mostly occur at a cognitive level and cannot result only from sensory priming. Bharucha and Stoeckig (1987) reported priming effects even when prime and target chords did not share any component notes. Tekman and Bharucha (1992) reported priming effects even when prime and target were separated by long silent intervals, and when white noise was introduced between prime and target. Moreover, in a recent study, we observed that harmonic relatedness resulted in a stronger priming effect than chord repetition (Bigand et al., in press). In the harmonic priming condition, the target chord (say a C major chord) was preceded by a musically highly related prime chord (a G major chord in this case). In the repetition priming condition, prime and target chords were identical (a C major chord followed by a C major chord). Repetition priming involves a strong

component of sensory priming since the two chords are identical. Harmonic priming involves strong top-down influences since the harmonic relation between prime and target corresponds to the most significant musical relationship in Western tonal music (i.e., an authentic cadence, which is a harmonic marker of phrase endings). In a set of five experiments, we never observed stronger priming effects in the repetition condition. Moreover, significantly stronger priming was observed in the harmonic priming condition in most of the experiments. This finding raises considerable difficulties for sensory models of music perception as the processing of a musical event is more facilitated when it is preceded by a different, but musically related chord than when it is preceded by an identical (repeated) chord.

These studies suggest that a single prime chord manages to activate an abstract knowledge of Western harmonic hierarchies. This activation results in the expectation that harmonically related chords should occur next. The present interpretation does not imply that sensory priming never affects chord processing. Indeed, Tekman and Bharucha (1998) showed that cognitive priming failed to overrule sensory priming when stimulus-onset-asynchrony (SOA) between chords was as short as 50 ms. In this experiment, the authors contrasted two types of prime and target relationships. In one type of chord pair, the target shared one note with the prime (C and E major chords)<sup>8</sup> but shared no parent major key. The other type of pair represented the opposite situation with the target sharing no note with the prime (C and D major chords), but both sharing a parent key (i.e., the key of G major). Consequently, the first pair favors sensory priming, while the second pair favors cognitive priming. The authors demonstrated that the processing of the target chord was facilitated in the second pair only for SOAs longer than 50 ms. This outcome suggests that top-down influences need some time to be instilled, while sensory priming occurs very quickly.

The influence of longer musical contexts on the processing of target chords has been addressed in several ways. In Bigand and Pineau (1997), eight-chord sequences were used with the last chord defining the target. The harmonic function of the target chord was varied by manipulating the first six chords of the sequence (Fig. 9.7). In the strongly expected condition, the target chord acted as a tonic chord (I). In the less expected condition, the target acted as a subdominant chord (IV), which was musically congruent with the context, but less expected. To reduce sensory priming effects, the chord immediately preceding the target was identical in both conditions. For the purpose of the experimental task, the target chord was rendered acoustically dissonant in half of the trials by adding a note to the chord. As a consequence, 25% of the trials ended on a consonant tonic chord, 25% on a consonant subdominant chord, 25% on a dissonant tonic chord, and 25% on a dissonant subdominant chord. Participants were required to indicate as accurately and as quickly as possible

---

<sup>8</sup>The major chords C, D, and E consist of the tones (C–E–G), (D, F#–A) and (E–G#–B), respectively.

Highly expected condition

Weakly expected condition

Target

Highly expected condition

Target

Weakly expected condition

Moderately expected condition

Dmajor I  
key

whether the target chord was acoustically consonant or dissonant. The critical finding of the study was to show that this consonant/dissonant judgment was more accurate and faster when targets acted as a tonic rather than as a subdominant chord. This suggests that the processing of harmonic spectra is facilitated for events that are the most predictable in the current context. Moreover, this study provided further evidence that musical expectancy does not occur from chord to chord, but also involves higher levels of musical relations.

This last issue was further investigated in Bigand et al. (1999) by using 14-chord sequences. As illustrated in Figure 9.7b, these chord sequences were organized into two groups of seven chords. The first two conditions replicated the conditions of Bigand and Pineau (1997) with longer sequences: chord sequences ended on either a highly expected tonic target chord or a weakly expected subdominant target chord. The third condition was new for this study and created a moderately expected condition. This third group of sequences was made out of the sequences in the first two conditions: The first part of the highly expected sequences (chords 1 to 7) defined the first part of this new sequence type and the second part of the weakly expected sequences (chords 8 to 14) defined their second part. The critical comparison was to assess whether the processing of the target chord is easier and faster in the moderately expected condition than in the weakly expected condition. This facilitation would indicate that the processing of a target chord has been primed in this third sequence by the very beginning of the sequence (the first seven chords which are highly related). The behavioral data confirmed this prediction. For both musician and nonmusician listeners, the processing of the target was most facilitated in the highly expected condition, followed by the moderately expected condition and then by the weakly expected condition. This finding further suggests that context effects can occur over longer time spans and at several hierarchical levels of the musical structure (see also Tillmann et al. 1998).

The effect of large musical contexts on chord processing has been replicated with different tasks. For example, in Bigand et al. (2001), chord sequences were played with a synthesized singing voice. The succession of the synthetic phonemes did not form a meaningful, linguistic phrase (e.g., /da fei ku jo fa to kei/). The last phoneme was either the phoneme /di/ or /du/. The harmonic relation of the target chord was manipulated so that the target acted either as a tonic or as a subdominant chord. The experimental session thus consisted of



FIGURE 9.7. (*Top*) One example of the eight-chord sequence used by Bigand and Pineau (1997) for the highly expected condition ending on the tonic chord (I) and the weakly expected condition ending on the subdominant chord (IV). From Bigand et al. (1999), Figure 1. Adapted with permission of the American Psychological Association. (*Bottom*) An example of the 14-chord sequences in the highly expected condition, the weakly expected condition and the moderately expected condition. From Bigand et al. (1999), Figure 6. Adapted with permission of the American Psychological Association.

50% of the sequences ending on a tonic chord (25% being sung with the phoneme *di*, 25% with the phoneme *du*) and 50% of sequences ending with a subdominant chord (25% sung with the phoneme *di*, 25% with the phoneme *du*). Participants performed a phoneme-monitoring task by identifying as quickly as possible whether the last chord was sung with the phoneme *di* or *du*. Phoneme-monitoring was shown to be more accurate and faster when the phoneme was sung on the tonic chord than on the subdominant chord. This finding suggests that the musical context is processed in an automatic way—even when the experimental task does not require paying attention to the music. As a result, the musical context induces auditory expectations that influence the processing of phonemes. Interestingly, these musical context effects on phoneme monitoring were observed for both musically trained and untrained adults (with no significant difference between these groups), and have recently been replicated with 6-year-old children. The influence of musical contexts was replicated when participants were required to quickly process the musical timbre of the target (Tillmann et al., 2004) or the onset asynchrony of notes in the target (Tillmann and Bharucha 2002).

These experiments differ from those run by Bharucha and collaborators not only by the length of the musical prime context, but also because complex musical sounds were used as stimuli (e.g., piano-like sounds in Bigand et al. 1999; singing voice-like sounds in Bigand et al. 2001) instead of Shepard notes. Given that musical sounds have more complex harmonic spectra than do Shepard notes, sensory priming effects should have been more active in the studies by Bigand and collaborators. A recent experiment was designed to contrast the strength of sensory and cognitive priming in long musical contexts (Bigand et al. 2003). Eight-chord sequences were presented to participants who were required to make a fast and accurate consonant/dissonant judgment on the last chord (the target). For the purpose of the experiment, the target chord was rendered acoustically dissonant in half of the trials by adding an out-of-key note. As in Bigand and Pineau (1997), the harmonic function of the target in the prime context was varied so that the target was always musically congruent: in one condition (highly expected condition), the target acted as the most referential chord of the key (the tonic chord) while in the other (weakly expected condition) it acted as a less referential subdominant chord. The critical new point was to simultaneously manipulate the frequency of occurrence of the target in the prime context. In the *no-target-in-context condition*, the target chords (tonic, subdominant) never occurred in the prime context. In this case, the contribution of sensory priming was likely to be neutralized. As a consequence, a facilitation of the target in the highly expected condition over the weakly expected condition could be attributed to the influence of knowledge-driven processes. In the *subdominant-target-condition*, we attempted to boost the strength of sensory priming by increasing the frequency of occurrence of the subdominant chord only in the prime context (the tonic chord never occurred in the context). In this condition, sensory priming was thus expected to be stronger, which should result in facilitated processing for subdominant targets.



In experiment 1, the consonant/dissonant task was performed more easily and quickly for tonic targets, and there was no effect of the frequency of occurrence. This finding suggests that top-down processes (cognitive priming) are more influential than sensory-driven process (sensory priming) in large musical contexts even though complex piano-like sounds were used. In experiment 2, the same sequences were used, but the tempo at which the sequences were played was increased. The slowest tempo was two times faster than in Experiment 1 (i.e., 300 ms per chord) and the highest tempo was 8 times faster (i.e., 75 ms per chord). The tempo variable was manipulated in blocks, with half of the participants starting the experiment with the slowest tempo and ending with the fastest tempo (group Slow–Fast). The other half of the participants started with the fastest tempo and ended with the slowest tempo (group Fast–Slow). On the basis of Tekman and Bharucha (1998), we expected that sensory priming would become more influential than cognitive priming with increasing tempo.

Our findings globally confirmed this hypothesis, with an interesting data pattern. At tempi of 300 ms and 150 ms per chord, priming effects were always stronger for tonic chords, irrespective of the target's frequency of occurrence. This data pattern changed at the fastest tempo (75 ms per chord), and there was a significant interaction with the temporal order at which the tempi were presented in the experimental session (i.e., groups Fast–Slow versus Slow–Fast). At this extremely fast tempo, sensory priming overruled cognitive priming only in the Fast–Slow group, and cognitive priming continued to be more influential in the Slow–Fast group. This second experiment sheds new light on the working of top-down processes in music by demonstrating that these processes continue to be more influential than sensory-driven processes even at a tempo as fast as 150 ms per chord.

This outcome highlights the speed at which the cognitive system manages to process abstract information (e.g., the musical function of a chord). At the tempo of 75 ms, sensory-driven processes overrule cognitive processes only in listeners who started to process musical sequences presented at this extremely fast tempo. The fact that cognitive priming continued to be more influential than sensory priming in the Slow–Fast group suggests that, once activated, the cognitive component continues to overrule sensory priming even at this extremely fast tempo. Once again, this complex pattern of data was observed for both musically trained and untrained listeners. This finding demonstrates that the auditory perception of musically untrained listeners is more sophisticated than generally assumed, at least for tasks involving the processing of complex pitch structures (e.g., musical chords). The weak difference observed in most of the studies cited above suggests that context effects in music involve robust, cognitive mechanisms.

## 5. Neurophysiological Bases of Context Effects in the Music Domain

Neurophysiological studies investigate the functioning of top-down processes by analyzing event-related potentials (ERPs) following contextually unexpected events, and by describing the cortical areas involved in these processes with the help of imaging techniques such as functional magnetic resonance imaging (fMRI). Different techniques allow the analysis of different aspects of the neurophysiological bases due to their inherent methodological advantages and limitations, which are notably linked to their temporal and spatial resolution. While electrophysiological methods, which are based on direct mapping of transient brain electric dipoles generated by neuronal depolarization (electroencephalography [EEG]) and the associated magnetic dipoles (magnetoencephalography [MEG]), provide fine temporal resolution of the recorded signal without precise spatial resolution, fMRI and positron emission tomography (PET) provide increased anatomical resolution of the implied brain structures, but the length of the measured temporal sample is rather long. Griffiths (Chapter 5) describes how these methods allow further understanding of processes linked to different pitch attributes and low-level perceptual processes. The present section focuses on the contribution of these techniques to our understanding of higher-level cognitive processes involved in auditory perception.

Numerous neurophysiological studies investigating top-down processes have used linguistic stimuli and visual stimuli (for a recent review of functional neuroimaging in cognition see Cabeza and Kingstone 2001). For context effects in language perception, evoked potentials following semantic and syntactic violations have been distinguished. At the end of a sentence (e.g., “The pizza was too hot to . . .”), the processing of a semantically unexpected word (e.g., “cry”) in comparison to an expected word (e.g., “eat”) evokes an N400 component (i.e., a negative evoked potential with a maximum amplitude 400 ms after the onset of the target word; Kutas and Hillyard 1980). By contrast, a syntactically incorrect sentence construction evokes a late positive potential (with a maximum amplitude 600 ms after the onset of the target word defining a P600 component) that has a larger amplitude than the potential evoked by a complex, but correct sentence structure (Patel et al. 1998). Moreover, in simple syntactic sentences, no P600 was observed. This outcome suggests that the amplitude of the P600 is inversely related to the ease of integrating a word into the previous context, with complex syntax and syntactic violation having a cost in terms of structural integration processes.

Over the last few years, a growing number of studies have used musical stimuli (e.g., Besson and Faïta 1995; Janata 1995; Koelsch et al. 2000; Regnault et al. 2001). Interestingly, the influence of a musical context has been shown to be associated with similar electrophysiological reactions as those observed in language perception: a given musical event evokes a stronger P300 (i.e., a positive evoked potential with a maximum amplitude 300 ms after the onset of the

target) or a late positive component (LPC, peaking around 500 and 600 ms) when it is unrelated to the context than when it is related. Besson and Faïta (1995) used familiar and unfamiliar melodies ending on either a congruous diatonic note,<sup>9</sup> an incongruous diatonic note or a nondiatonic note. At the onset of the last note of the melodies, the amplitude of the LPC component was stronger for the nondiatonic note than for the incongruous diatonic ones and the weakest for the congruous diatonic notes. Other studies have analyzed the event-related potentials consecutive to a violation of harmonic expectancies (i.e., for chords). Consistent with Besson and Faïta (1995), it was shown that the amplitude of the LPC increases with increasing harmonic violation: the positivity was larger for distant-key chords than for closely related or in-key chords (Janata 1995; Patel et al. 1998). In Patel et al. (1998), for example, target chords that varied in the degree of their harmonic relatedness to the context occurred in the middle of musical sequences: the target chord may be the tonic chord of the established context key or may belong to a closely related key, or it may belong to a distant, unrelated key. The target evoked an LPC with largest amplitude for distant-key targets, and with decreasing amplitude for closely related key targets and tonic targets. Patel et al. (1998) compared directly the evoked potentials due to syntactic relationships and harmonic relations in the same listeners: both types of violations evoked an LPC component, suggesting that a late positive evoked potential is not specific to language processing, but reflects more general structural integration processes based on listeners' knowledge.

The neurophysiological correlates of musical context effects are reported also for finer harmonic differences between target chords. Based on the priming material of Bigand and Pineau (1997), Regnault et al. (2001) attempted to separate two levels of expectations—one linked to the context (related versus less-related targets) and one linked to the acoustic features of the target in the harmonic priming situation (consonant versus dissonant targets). Related targets and less-related targets correspond to the tonic and subdominant chords represented in Figure 9.6. In half of the trials, these targets were rendered acoustically dissonant by adding an out-of-key note in the chord (e.g., a C# to a C major chord). The experimental design allows an assessment of whether violations of cognitive and sensory expectancies are associated with different components in the event-related potentials. For both musician and nonmusician listeners, the violation of cognitive and sensory expectancy was shown to result in an increased positivity at different time scales. The less-related, weakly expected target chords (i.e., subdominant chords) evoked a P3 component (200 to 300 ms latency range) with larger amplitude than that of the P3 component linked to strongly related tonic targets. The dissonant targets elicited an LPC component (300 to 800 ms latency range) with larger amplitude than the LPC of consonant targets. This outcome suggests that violations of top-down expectancies are detected very quickly, and even faster than violations of sensory dissonance. The observed fast-acting, top-down component is consistent with

---

<sup>9</sup>Diatonic notes correspond to notes that belong to the key context.

behavioral measures reported in a recent study designed to trace the time course of both top-down and bottom-up processes in long musical contexts (Bigand et al. 2003, and see Section 4). In addition, the two components (P3, LPC) were independent; notably the difference in P3 amplitude between related and less-related targets was not influenced by the acoustic consonance/dissonance of the target. This outcome suggests that musical expectancies are influenced by two separate processes. Once again, this data pattern was reported for both musically trained and untrained listeners: both groups were sensitive to changes in harmonic function of the target chord due to the established harmonic context.

Nonmusicians' sensitivity to violations of musical expectancies in chord sequences has been further shown with ERPs (Koelsch et al. 2000) and MEG (Maess et al. 2001) for the same harmonic material. In the ERP study, an early right-anterior negativity (named ERAN, maximal around 150 ms after target onset) reflected the harmonic expectancy violation in the tonal contexts. The ERAN was observed independently of the experimental task: for example, the detection of timbral deviances while ignoring harmonies (experiments 1 and 2) or the explicit detection of chord structures (experiments 3 and 4). Unexpected events elicited both an ERAN and a late bilateral frontal negativity, N5 (maximal around 500 to 550 ms). This latter ERP component N5 was interpreted in connection with musical integration processes: its amplitude decreased with increasing length of context and increased for unexpected events. A right-hemisphere negativity (N350) in response to out-of-key target chords has been also reported by Patel et al. (1998, right antero-temporal negativity, RATN) who suggested links between the RATN and the right fronto-temporal circuits that have been implicated in working memory for tonal material (Zatorre et al. 1994). It has been further suggested by Patel et al. (1998) and Koelsch et al. (2000) that the right early frontal negativities might be related to the processing of syntactic-like musical structures. They compared this negativity with the left early frontal negativity ELAN observed in auditory language studies for syntactic incongruities (e.g., Friederici 1995; Friederici et al. 2000). This component is thought to arise in the inferior frontal regions around Broca's area.

The implication of the prefrontal cortex has also been reported for the manipulation and evaluation of tonal material, notably for expectancy violation and working memory tasks (Zatorre et al. 1992, 1994; Patel et al. 1998; Koelsch et al. 2000). Further converging evidence for the implication of the inferior frontal cortices in musical context effects has been provided by Maess et al.'s (2001) study using magneto-encephalography measurements on the musical sequences of Koelsch et al. The deviant musical events evoked an increased bilateral mERAN (the magnetic equivalent of the ERAN) with a slight asymmetry to the right for some of the participants. The generators of this MEG signal were localized in Broca's area and its right hemisphere homologue. Koelsch et al. (2002) investigated with fMRI the neural correlates of musical sequences similar to previously used material (Koelsch et al. 2000; Maess et al. 2001): chord sequences contained infrequently presented unexpected musical events. The observed activation patterns confirmed the implication of Broca's area (and ante-

rior–superior insular cortices) in the processing of musical violations. The reported network further included Wernicke's area as well as superior temporal sulcus, Heschl's gyrus and both planum polare and planum temporale.

A recent fMRI study investigated neural correlates of target chord processing in a musical priming paradigm (Tillmann et al. 2003). In eight-chord sequences, the last chord defined the target that was either strongly related (a tonic chord) or unrelated (a chord belonging to a different, unrelated key). As in previous musical priming studies, half of the targets were rendered acoustically dissonant for the experimental task. Participants were scanned with fMRI while performing speeded intonation judgments (consonant versus dissonant) on the target chords. Behavioral results acquired in the scanner replicated the facilitation effect of related over unrelated consonant targets. The overall activation pattern associated with target processing showed commonalities with networks previously described for target detection and novelty processing (Linden et al. 1999; Kiehl et al. 2001). This network included activation in frontal areas (inferior, middle and superior frontal gyri, insula, anterior cingulate) and posterior areas (inferior parietal gyri, posterior cingulate) as well as in the thalamic nuclei and the cerebellum. The characteristics of the targets, notably in how far the chord fit or violated the expectations built up by the prime context, influenced the activation levels of some of these network components. Increased activation was observed for targets that violated expectations based on either sensory–acoustic or harmonic relations. For example, the activation in bilateral inferior frontal regions (i.e., inferior frontal gyrus, frontal operculum, insula) was stronger for unrelated than for related (consonant) targets. The strength of activation in these areas also indicated the detection of dissonant targets in comparison to consonant targets.

The manipulation of harmonic relationships in this fMRI study was extremely strong: in the related condition, the target played the role of the most important, stable chord (i.e., the tonic) and in the unrelated condition the target did not even belong to the key of the prime context. Consequently, the two targets had either strong or weak association strengths to the other chords of the prime context. When analyzing musical pieces of the Western tonal repertoire, it will become evident that the related target chord is frequently associated with chords of the prime context, while the unrelated target chord is not. The musical priming study reported increased activation in (bilateral) inferior frontal areas for targets weakly associated to the prime events (the unrelated targets). Interestingly, language studies that manipulated associative strengths between words also reported increased inferior frontal activation for weakly associated words (Wagner et al. 2001) or semantically unrelated word pairs (West et al. 2000). The strong manipulation of the harmonic relationships has a second consequence: the notes of the related target occurred in the prime context while the notes of the unrelated target did not. In other words, in these musical sequences sensory and cognitive priming worked in the same direction and favored the related target. It is interesting to make the link with other functional imaging data reporting the phenomenon of repetitive priming for the processing of ob-

jects and words: decreased inferior frontal activation is observed for repeated items in comparison to novel items (Koustaal et al. 2001). This finding suggests that weaker activation for musically related targets might also involve repetition priming for neural correlates in musical priming. This hypothesis, which needs further investigation, is very challenging as behavioral studies (reported above) provide evidence for strong cognitive priming (Bigand et al. 2003).

The outcome of the musical priming study is convergent with Maess's source localization of the MEG signal after a musical expectancy violation. The present data sets on musical context effects can be integrated with other data showing that Broca's area and its right homologue participate in nonlinguistic processes (Pugh et al. 1996; Griffiths et al. 1999; Linden et al. 1999; Müller et al. 2001; Adams and Janata 2002) besides their roles in semantic (Poldrack et al. 1999; Wagner et al. 2000), syntactic (Caplan et al. 1999; Embick et al. 2000), and phonological functions (Pugh et al. 1996; Fiez et al. 1999; Poldrack et al. 1999). Together with the musical data, current findings point to a role of inferior frontal regions for the integration of information over time (cf. Fuster 2001). The integrative role includes storing previously heard information (e.g., a working memory component) and comparing the stored information with further incoming events. Depending on the context, listener's long-term memory knowledge about possible relationships and their frequencies of occurrence (and co-occurrence) allows the development of expectations for typical future events. The comparison of expected versus incoming events allows the detection of a potential deviant and incoherent event. The processing of deviants, or more generally of less frequently encountered events, may then require more neural resources than processing of more familiar or prototypical stimuli.

## 6. Implicit Learning of Pitch Regularities

One finding reported in most of the studies described above may have surprised the reader. Top-down influences on perception, memorization, and processing of pitch structures were consistently shown to depend only weakly on the extent of musical expertise. This finding contradicts the common belief that musical experts should perceive music differently than musically untrained (supposedly naive) listeners. In the reported experimental studies, musically untrained listeners are sensitive to the same contextual factors as musician listeners, and these factors influence perceptual behavior (and neurophysiological correlates) in roughly the same way as for musician listeners. This outcome suggests that top-down processes are acquired through robust processes that do not require explicit training. This conclusion raises an intriguing question: How can the pitch structure regularities of our environment be internalized by the human brain? In this section, we argue that implicit learning processes that have been investigated in several domains in cognitive psychology are likely to occur as well in the auditory domain and particularly in the music domain. Section 7 then proposes how these processes might be formalized in a neural net model.

Implicit learning describes a form of learning in which subjects become sensitive to the structure of a complex environment through simple, passive exposure to that environment. Reber (1992) considers this type of learning to be a fundamental cognitive process that permits the acquisition of complex information, which is inaccessible to deductive reasoning. Implicit learning has some specific characteristics that distinguish it from explicit learning processes: implicitly acquired knowledge remains longer in memory (Allen and Reber 1980), is less sensitive to interindividual differences (Reber et al. 1991), and is more resistant to cognitive and neurological disorders (Abrams and Reber 1988).

The most famous experimental protocols to study implicit learning consist of presenting participants with sequences of events (e.g., letters, light positions, sounds) generated by an artificially defined grammar. Figure 9.8 displays a sample grammar similar to the grammar first used by Reber (1967, 1989). The arrows represent legal transitions between the different letters (X–S–J–Q–W), and a loop indicates possible repetitions of a letter (X or S in this case). During the first phase of the experiment, participants were exposed to sequences of letters that conform to the rules of the grammar (e.g., WJSSX; XSWJSX). One group of participants was asked to discover the rules that generate the grammar (Explicit Condition), while the other group was asked to memorize the sequences and was unaware that any rules existed (Implicit Condition). During the second phase of the experiment, the participants were informed that the sequences of the first phase had been produced by a rule system (which was not described to them). The participants were then asked to judge the grammaticality of new letter sequences. Half of these sequences were ungrammatical (e.g., XSQJ, WSQX) and half were new grammatical exemplars. In general, participants in the Implicit Condition performed better than those in the Explicit Condition (varying between 60% and 80% of correct responses). Only a few participants of the implicit group were able to describe aspects of the rules used to generate the letter sequences. As stated initially by Reber (1967, 1989), participants acquired an implicit knowledge of the abstract rules of the grammar.

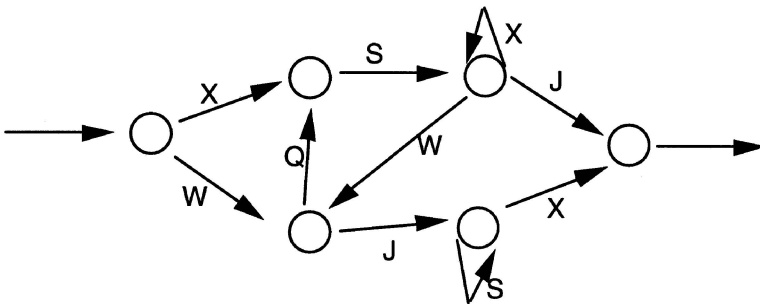


FIGURE 9.8. Example of a finite state grammar generating letter sequences. The sequence XSXXWJX is grammatical whereas the sequence XSQSW is not.

The very nature of the knowledge acquired in these experimental situations, as well as the complete implicit nature of this knowledge has been a matter of debate and still is now (see Perruchet and Pacteau 1990; Perruchet et al. 1997; Perruchet and Vinter 2002), but it is largely admitted that passive exposure results in the internalization of regularities underlying the variations of the external environment.

Although auditory stimuli were rarely used in the domain of implicit learning, some empirical findings demonstrate that regular structures of the auditory environment can also be internalized through passive exposure. A strict adaptation of Reber's study to the auditory domain was realized by Bigand et al. (1998), with letters being replaced by musical sounds of different timbres (e.g., gong, trumpet, piano, violin, voice). In the first phase of the experiment, participants listened to sequences of timbres that obeyed the rules of an artificial grammar. The Implicit group was asked to memorize the sequences and to indicate whether a particular timbre sequence was heard for the first or the second time. The Explicit group was required to memorize the timbre sequences and was told that these sequences had been produced by a computer program. Participants of this group were encouraged to try to identify these rules and were told that discovering these rules would contribute to better memory performance. After this first exposition phase, both groups were required to differentiate grammatical and ungrammatical sequences of timbres. A control group was added that performed this last phase without having been exposed to the grammatical sequences. Explicit and Implicit groups performed better than the control group in the grammatical task, with the performance of the Implicit group being slightly better than that of the Explicit group. This outcome suggests that prior exposure to a small number of timbre sequences governed by an artificial rule system was sufficient to enable participants to determine the new sequences that broke one or more of these rules. The internalization of the timbre grammars may therefore result from the simple exposure to sequences generated by the system without the necessity to implement any explicit process of analysis.

A very elegant demonstration of the strength of implicit learning in the auditory domain was provided by Saffran and collaborators. In their initial experiments (Saffran et al. 1996, 1997), meaningless phonemes were presented to adults, children, and infants in a continuous sequence (e.g., bupadapatubitutibu...). The phoneme sequence was constructed with several artificial three-syllable words (e.g., bupada, patubi) chained together without pauses or other surface cues. Consequently, the transition probabilities between two syllables<sup>10</sup> allowed finding word boundaries: transition probabilities inside a word were high, but transition probabilities across word boundaries were weak. If listeners became sensitive to these statistical regularities, they would be able to extract the words from this artificial language. The experiments consisted of two phases. In a first exposition phase, participants listened to the continuous stream

---

<sup>10</sup>The transition probability that A is followed by B is defined by the frequency of the pair AB divided by the frequency of A (Saffran et al. 1996).



for about 20 minutes (Saffran et al. 1996 for adults) while performing either a coloration task or doing nothing. In the second phase of the experiment, participants were tested with a two-alternative forced-choice task: a real word of the artificial language and a nonword (three syllables that do not create a word) were presented in pairs, and participants had to indicate which one belonged to the previously heard sequence. Participants performed above chance in this task, even when words were contrasted to so called *part-words* in which two syllables were part of a real word, but the association with the third syllable was illegal.<sup>11</sup> In infant experiments, the testing phase was based on novelty preferences (and the dishabituation effect): infants' looking times were longer for the loudspeaker emitting nonwords than for the loudspeaker emitting words. The simple exposure to the sequence of phonemes results in the internalization of artificial words even for 8-month-old infants. With the goal to show that the capacity to extract these statistical regularities is not restricted to linguistic material, Saffran et al. (1999) replaced the syllables by pure tones in order to create *words of tones*, which, once again, are concatenated continuously to each other to create a sequence. The tones were carefully chosen in such a way that the tone words and the chaining of these words in the sequence did not create a specific key context, and overall, they did not respect tonal rules nor did they resemble familiar three-tone sequences (e.g., the NBC television network's chimes). After exposition, both adults and 8-month-old infants performed above chance in the testing phase and performed as well as for linguistic-like sequences of syllables. Listeners thus succeeded in segmenting the tone stream and in extracting the *tone units*. Overall, Saffran et al.'s data suggest that statistical learning of different materials can be based on similar knowledge-acquisition processes.

To some extent, this finding can be considered as illustrating in the laboratory the processes that actually occur in real life for extensive exposure to environmental sounds, including music. It is obvious that a musical system such as the Western tonal system is more complex than the artificial grammar exposed in Figure 9.8. However, the opportunities to be exposed to sequences obeying this system from birth (and probably 3 or 4 months before birth) are so numerous that most of the rules of Western tonal music may be internalized through similar processes. Following this hypothesis, Western listeners may have acquired a sophisticated knowledge about Western tonal music, even though this knowledge remains at an implicit level of representation. A large set of empirical studies has actually demonstrated that musically untrained listeners (even young children) have internalized several aspects of the statistical regularities underlying pitch combinations that are specific to Western tonal music (Francès 1958; Thompson and Cuddy 1989; Krumhansl 1990; Cuddy and Thompson 1992a,b; see Bigand 1993 for a review). Some extensions to other musical cultures have been realized in single studies (Castellano et al. 1984; Krumhansl et al. 1999).

---

<sup>11</sup>For example, for the word "bupada" a part-word would contain the first two syllables followed by a third different syllable "bupaka" (with the constraint that this association does not form another word).

Once acquired, this implicit knowledge induces fast and rather automatic top-down influences on the perception and processing of Western pitch structures and renders musically untrained listeners “musically expert” for the processing of these pitch structures. One critical issue that remains is to formalize the functioning of these implicit learning processes in the auditory domain. The last section provides some first insights into this issue.

## 7. Neural Net Modeling of Implicit Learning of Western Pitch Structures

Pitch models and models of basic processes of pitch perception have been presented by de Cheveigné (Chapter 6). The present section focuses on models of music perception, and particularly artificial neural networks that simulate the learning and perceiving of musical structures. One of the principal advantages of artificial neural networks (e.g., connectionist models) is their capacity to learn representations, categorizations, or associations between events. In these networks, the rules governing the material are not stored in an explicit (symbolic) way, but emerge from multiple constraints represented by the connections of the network, which have been learned by repeated exposure. In the following, some basics of neural net modeling will be reviewed first, followed by applications of neural nets to music perception. In this line, a model using self-organizing maps (SOMs) will be presented as one example of neural nets simulating the learning and perception of musical structures.

An artificial neural network consists of units linked via synaptic connections of different strengths. The units are generally arranged into layers, with an input layer coding the incoming information. The input units are activated when a stimulus is presented to the network. This activation is sent via the connections to units in other layers. The strength of the transmitted activation is determined by the strengths of the connections (i.e., weights of the connections). At the outset, a network does not incorporate any knowledge of the material, and this ignorance is reflected by connection weights set to random values. In parallel with biological networks, the learning process is defined as a modification of connection weights (Hebb 1949). Over the course of learning, the neural net units gradually become sensitive to different input events or categories. The learning process can be either supervised by an external teaching exemplar (e.g., the delta rule, McClelland and Rumelhart 1986) or unsupervised via passive exposure (e.g., competitive learning, Rumelhart and Zipser 1985). In supervised learning algorithms, an external teaching instance prescribes the target output that has to be reached and the weights of the connections are modified so that the model’s output matches this target. In unsupervised learning algorithms, the network adapts its connections in such a way that it becomes sensitive to the underlying correlational structure between events of the training set: statistical regularities of the input material are extracted and events that often occur together are encoded and represented by the net units. As acculturation to musical

structures presumably occurs without supervision in listeners, unsupervised learning algorithms seem to be well suited to modeling music cognition. The present section thus focuses on unsupervised learning algorithms, notably the competitive learning algorithm that provides the basis for learning in SOMs (Kohonen 1995) and artificial resonance theory (ART) networks (see Grossberg 1970, 1976).

For the competitive learning process, a set of training stimuli is presented repeatedly to the network and the learning takes place by competition among the units (Rumelhart and Zipser 1985). When an input is presented to the network, the input layer sends activation via the random connection weights to the units of the next layer. The unit receiving the maximum activation is defined as the “winner” of the competition (e.g., best representing the current input) and is allowed to learn the representation of this input even better. Following the learning rule, the weights of the connections are updated in such a way that the links coming from active input units are reinforced and links coming from inactive input units are weakened. In other words, the response of the winning unit will subsequently be stronger for this same input pattern (or similar ones) and weaker for other patterns. In a similar way, other units learn to specialize their responses to other input patterns. The competitive learning algorithm represents the basis for learning in SOMs. In a network using an SOM, the units that are connected to the input layer follow a spatial layout: units are arranged in the form of a map and neighborhood relationships can be defined between map units as a function of the distance between these units. For learning in an SOM, not only the winning unit, but also the neighboring units are allowed to learn. At the beginning of learning, the size of the neighborhood is broad and over the course of learning its radius decreases. This learning process leads to topological mappings between input data and neural net units on the map: units that respond maximally for similar input patterns are located near each other on the map. Topological organization conforms to principles of cortical information processing, such as spatial ordering in sensory processing areas (e.g., somatosensory, vision, audition). Winter (Chapter 4) and Griffiths (Chapter 5) review the tonotopic organization of the auditory system that can be found at almost all major stages of processing (i.e., inner ear, auditory nerve, cochlear nucleus and auditory cortex).

Neural nets based on unsupervised learning algorithms are helpful in understanding how we learn musical patterns by mere exposure, how these patterns might be represented, and how this knowledge arising from acculturation influences perception. Recently, we used the SOM algorithm to simulate the cognitive capacity to extract underlying regularities and to become sensitive to musical structures via implicit and unsupervised learning processes (Tillmann et al. 2000). Western tonal musical pieces are based on a three-level organizational system containing notes, chords, and keys (cf. Section 2). For the simulation of the implicit learning of tonal regularities, a hierarchical network with two SOMs was defined. The units of the input layer coded the incoming 12 pitch classes, taking into consideration octave equivalence. Each unit of the input

layer was connected to the units of the first SOM that in turn were connected to the units of the second SOM. Before learning, the weights of all connections were set to random values. During learning, chords and chord sequences were presented repeatedly to the input layer of the network. The connectionist algorithm changed connections in order to allow units to become specific detectors of combinations of events over short temporal windows. The structure of the system adapted to the regularities of tonal relationships through repeated exposure to musical material. Over the course of learning, the weights of the connections changed to reflect the regularities of co-occurrences between notes and between chords. The first connection matrix reflects which pitch (or virtual pitch) is part of a chord; the second matrix reflects which chord is part of a key. The units of the first SOM became specialized for the detection of chords and the units of the second SOM for the detection of keys. Both SOM layers showed a topological organization of the specialized units. In the chord layer, units representing chords that share notes (or subharmonics) were located close to each other on the map, but chords not sharing notes were not represented by neighboring units. In the key layer, the units specialized in the detection of keys were organized in a circle: keys sharing numerous chords and notes were represented close to each other on the map and the distance between keys increased with decreasing number of shared events. The organization of key units reflects the music theoretic organization of the circle of fifths: the more the keys are harmonically related, the closer they are on the circle (and on the network map). The learnability of this kind of higher-level topological map (cf. also Leman 1995) has led to the search for neural correlates of key maps (Janata et al. 2002).

The hierarchical SOM thus managed to learn Western pitch regularities via mere exposure. The entire learning process is guided by bottom-up information only and takes place without an external teacher. Furthermore, there are no explicit rules or concepts stored in the model. The connections between the three layers extract via mere exposure how the events appear together in music. The overall pattern of connections reflects how notes, chords, and keys are interrelated. Just as for nonmusician listeners, the tonal knowledge is acquired without explicit instruction or external control. The input layer of the present network was based on units coding octave equivalent pitch classes. This model can be conceived as being on the top of other networks that have learned to extract pitch height from frequency (Sano and Jenkins 1991; Taylor and Greenhough 1994; Cohen et al. 1995) and octave-equivalent pitch classes from spectral representations of notes (Bharucha and Mencl 1996).

The SOM model integrates three levels of organization of the musical system. Other neural net models have been proposed in the literature that focused on either one or two organizational levels of music perception as, for example, pitch perception (Sano and Jenkins 1991; Taylor and Greenhough 1994), chord classification (Laden and Keefe 1991), or melodic sequence learning (Bharucha and Olney 1989; Page 1994; Krumhansl et al. 1999). More complex aspects of musical learning that are linked to the perception of musical style have been

simulated by Gjerdingen (1990) using an ART network. Other models focused more strongly on the preprocessing of the auditory signal by auditory modules and on the bottom-up processes involved in learning and perception (Leman 1995, 2000; Leman and Carreras 1998).

As presented up to this point, one characteristic of neural networks is the adaptation to environmental structures and the learning of a representation of the regularities inherent in the environment. Another attractive characteristic of neural networks is the possibility of accounting for top-down influences and for the way they combine with bottom-up influences. In the language domain, neural net models of word recognition (McClelland and Rumelhart 1981; Rumelhart and McClelland 1982) and of speech recognition (Elman and McClelland 1984; McClelland and Elman 1986) simulate the top-down influences of the knowledge representation via activation reverberating between layers, notably interactive activation between higher level units (words) and lower level units (letters or phonemes). When, for example, part of the written word is missing, the reverberating activation helps to select possible candidates and to restore information in order to recognize the word. In music perception, Bharucha (1987) proposed a model (referred to as MUSACT) that relies on a comparable architecture including a mechanism of spreading activation. In MUSACT, note units are connected to chord units that in turn are connected to key units. When a stimulus is presented to the model, note units are activated and activation reverberates in the system until equilibrium is reached. This reverberation mechanism simulates the top-down influences and changes the activation patterns in favor of culturally defined relationships. For example, when a C major chord (i.e., consisting of the notes C–E–G) is presented to the network, the activation pattern of the chord layer at the beginning reflects bottom-up influences only, notably the chord unit of E major will be more activated than the chord unit of D major because it shares one note with the stimulus chord (e.g., the note E), even if the chords C major and D major are harmonically more closely related. After reverberation, activation patterns change qualitatively and mirror theoretic Western harmonic hierarchies: the chord unit of D major now receives stronger activation than the chord unit of E major. The model thus predicts sensory priming for extremely short time spans with a facilitation of the E Major chord over the D major chord, and cognitive priming for longer time spans with a facilitation of the D major chord over the E Major chord. The model thus succeeds in simulating the time course of bottom-up and top-down activation as reported in short context priming by Tekman and Bharucha (1998, cf. Section 4). The MUSACT model has also simulated a set of priming data showing an effect of cognitive top-down influences in chord processing (Tillmann et al. 1998; Bigand et al. 1999, 2003; Tillmann and Bigand 2001; Tillmann et al. 2003).

However, MUSACT represents an idealized end-state of an implicit learning process as it is based on music theoretic constraints and neither connections nor weights resulted from a learning process. As reported earlier, a representation of pitch regularities (as implemented by MUSACT) can be learned by passive self-organization (cf. Tillmann et al. 2000). In addition to testing this learned

model with priming material (as was done with the MUSACT model), the SOM model has been tested for its capacity to simulate a variety of empirical data on the perceived relationships between and among notes, chords, and keys. For these simulations, the experimental material of behavioral studies was presented to the network and the activation levels of the network units were interpreted as levels of tonal stability. The more a unit (i.e., a chord unit or a note unit) is activated, the more stable the musical event is in the corresponding context. For the experimental tasks, it was hypothesized that the level of stability affects performance (e.g., a more strongly activated, stable event is more expected or judged to be more similar to a preceding event). The simulated data covered a range of experimental tasks, notably similarity judgments, recognition memory for notes and chords, priming, electrophysiological measures for chords, and perception and detection of modulations and distances between keys. Overall, the simulations showed that activation in the learned SOM model mirrored the data of human participants in a range of experiments on the perception of tonality (cf. Tillmann et al. 2000, for further details of individual results).

The SOM simulations provide an example of the application of artificial neural networks to increasing our understanding of learning and representing knowledge about the tonal system and the influence of this knowledge on perception and processing. The learning process can be simulated by passive exposure to musical material, just as it is supposed to happen in nonmusician listeners. Once acquired, the knowledge influences perception. It is worth underlining that the SOM model simulates a set of context effects linked to the perception of notes and of chords: the same chord unit is activated with different levels of activation depending on the tonality of the preceding context. For example, the model simulates the principles of contextual distance and contextual asymmetry observed for human participants in the similarity judgments of chord pairs presented above in Section 3 (Krumhansl et al. 1982a; Bharucha and Krumhansl 1983): the activation level of a chord unit changes as a function of the harmonic distance to the preceding key context and of the temporal order of presentation in the pair. The learned musical SOM network thus provides a low-dimensional and parsimonious representation of tonal knowledge: the contextual dependency of musical functions of an event emerges from the activation reverberating in the system, and the important stable events (e.g., musical prototypes and anchor points of a key) do not have to be stored separately in different units for each of the possible keys.

## 8. Conclusion

Throughout this chapter, we have documented that the processing of pitch structures is strongly context dependent. These context effects have been shown for the perception of specific attributes of musical sounds (such as musical stability), for the memorization of pitch (Section 3), as well as for the speed and accuracy of processing perceptual attributes related to the pitch dimension (e.g., sensory

dissonance, musical timbre, phoneme, Section 4). These top-down influences involve rather specific electrophysiological responses and cortical areas that, interestingly, seem not to differ radically between language and musical domains. This suggests that some brain structures may be specialized in the integration of contextual information. The ecological interest of this specialization might be, notably, to enhance the processing of the pitch dimension. Most of the examples reported to illustrate context effects come from the music domain (similar examples are, of course, numerous in spoken language). Probably, composers have intuitively developed a musical system that taps into this incredible flexibility of the auditory system to attribute perceptual sound qualities as a function of the context in which they appear. The Western musical system takes advantage of this fundamental feature of the human brain: the ability to interpret sensory input differently depending on the current context. The Western tonal system is remarkable from this point of view. Despite a very small number of pitch classes (12), an infinite number of musical sequences can be composed by taking advantage of context effects in perception and by modifying the perceptual qualities of musical sounds as a function of the current context.

Of course, the question arises as to whether this feature of context dependency is unique for Western tonal music or whether other musical systems use it. In Section 6, we argue that the observations made with musical material are just one example manifesting the broad competence of the human brain to internalize statistical regularities of environmental structures. Attempts to confirm implicit learning processes in the auditory domain have been presented with different sets of artificial materials. It is likely that new musical grammars will be internalized through passive exposure in roughly the same way. On the basis of this internalized knowledge, similar context effects will probably be reported in the future for contemporary music, as well as for other artificial sound structures that are derived from similar principles. Given the strength of the implication of implicit learning in auditory processing, we addressed in the previous section how these processes may be formalized in a neural net model.

We hope the present chapter will encourage new researchers to spend more time investigating the role played by implicit learning and top-down processes in the auditory domain. It is striking that learning and top-down processes are concepts that are missing in most current textbooks on audition (but see McAdams and Bigand 1993 for auditory cognition and SHAR textbooks for learning, plasticity and development—Rubel et al. 1997; Parks et al. 2004). Interestingly, audition is almost missing in the literature on implicit learning as well as on perceptual learning. As a consequence, the role of a listener's knowledge on auditory perception remains unclear and its importance is often disregarded or not even acknowledged. A better understanding of context effects in auditory perception has two possible main implications for the future. The first one is that adding knowledge and top-down processes in artificial models of auditory perception (including models of pitch processing) is likely to improve the models (see Carreras et al. 1999). There is strong evidence showing that the human brain manages to process pitch in a sophisticated way with the help

of these top-down processes. The way this knowledge is represented in the mind, as well as the way this knowledge is acquired through exposure needs to be documented in more detail. Our preliminary findings on pitch structures suggest that similar processes of learning may then be implemented in artificial systems so that they manage to simulate top-down processes (for a discussion of this issue in visual perception see Herzog and Fahle 2002).

The second main implication, which is only beginning to be considered, concerns the rehabilitation of hearing-impaired listeners. Over the last year, research projects are emerging that investigate learning processes in hearing-impaired listeners and patients with cochlear implants. However, up to now, this research mainly focuses on perceptual processes in audition, as, for example, loudness perception (Philibert et al. 2002), sound localization and binaural hearing cues (Moore 2002), or on phoneme processing and single-word processing without considering extended contexts (Clark 2002).

Numerous research has now generally established that top-down processes result in perceptual expectancies that enhance signal detection (e.g., Howard et al. 1984 for pitch detection threshold) and signal processing in all sensory modalities. Reinforcing these top-down processes in hearing-impaired listeners represents a key concern since the top-down processes should contribute to a compensation of the failure of sensory processes. Of course, this kind of strategy occurs naturally in hearing-impaired listeners and is usually developed by auditory teaching methods. However, it is obvious that the more the scientific community knows about the functioning of top-down processing as well as the functioning of learning processes in the auditory domain, the more efficient such teaching methods will be. Several factors that influence implicit auditory learning need to be studied in the auditory domain, and the benefits drawn from implicit versus explicit training should be evaluated. The outcome of this research will also have implications for research in technical engineering devoted to the remediation of hearing-impaired listeners. Up to now, considerable effort has been made for the investigation and the improvement of reception and coding of auditory signals at peripheral levels of processing. It is now necessary to invest in technical support favoring the development and improvement of higher-level, cognitive processes and perceptual top-down strategies that will then help the listener to restore missing or deficient sensory signals, to the extent that such is possible. During the last decade, cognitive engineering devoted to training techniques has been developed and strongly improved in several domains. These new technologies offer considerable possibilities to define auditory learning programs that will encourage implicit learning of auditory sound and scene structures. To take best advantage of these new technologies for hearing-impaired listeners, it is important that the scientific community involved in audition reinforces considerably the research programs on perceptual and statistical learning in audition.



## 9. Summary

This chapter focused on the effect of listeners' knowledge on the processing of pitch structures. In Section 2, several examples taken from vision and audition illustrated the differences between sensory processes and knowledge-driven processes (also referred to as bottom-up and top-down processes). Empirical evidence for top-down effects on the processing of pitch structures (perception and memorization) was presented in Sections 3 and 4. It has been shown that a long series of musical notes can be perceived differently as a function of the musical key context in which the notes occur, and that the speed and accuracy with which some qualities of musical chords (e.g., consonance versus dissonance, harmonic spectra) are processed depends on the musical function of the chord in the current context. The neurophysiological structures implied in top-down processes in music perception were reviewed in Section 5. Sections 6 and 7 addressed the origins of knowledge-driven processes. It was argued that a fundamental characteristic of the human brain is to internalize the statistical regularities of the external environment. In the case of music, intense passive exposure to Western musical pieces results in an implicit knowledge of Western musical regularities, which, in turn, govern the processing of pitch structures. The way implicit learning processes might be formalized by neural net models was developed in Section 7. In conclusion, it was emphasized that the context effects observed in music perception reflect the considerable importance of top-down processes in the auditory domain. This conclusion has several implications, notably for artificial models of pitch processing as well as for auditory training methods designed for hearing-impaired listeners.

## References

- Abrams M, Reber AS (1988) Implicit learning: robustness in the face of psychiatric disorders. *J Psycholing Res* 17:425–439.
- Adams RB, Janata P (2002) A comparison of neural circuits underlying auditory and visual object categorization. *NeuroImage* 16:361–377.
- Allen R, Reber AS (1980) Very long term memory for tacit knowledge. *Cognition* 8: 175–185.
- Ballas JA, Mullins T (1991) Effects of context on the identification of everyday sounds. *Hum Perform* 4:199–219.
- Bartlett JC, Dowling WJ (1980) The recognition of transposed melodies: a key-distance effect in developmental perspective. *J Exp Psychol Hum Percept Perform* 6:501–515.
- Bartlett JC, Dowling WJ (1988) Scale structure and similarity of melodies. *Music Percept* 5:285–314.
- Berlioz H (1872) *Mémoires*. Paris: Flammarion.
- Besson M, Faïta F (1995) An event-related potential (ERP) study of musical expectancy: comparison of musicians with nonmusicians. *J Exp Psychol Hum Percept Perform* 21:1278–1296.
- Bharucha JJ (1984a) Event hierarchies, tonal hierarchies, and assimilation: a reply to Deutsch and Dowling. *J Exp Psychol Gen* 113:421–425.

- Bharucha JJ (1984b) Anchoring effects in music: the resolution of dissonance. *Cognit Psychol* 16:485–518.
- Bharucha JJ (1987) Music cognition and perceptual facilitation: a connectionist framework. *Music Percept* 5:1–30.
- Bharucha JJ (1996) Melodic anchoring. *Music Percept* 13:383–400.
- Bharucha JJ, Krumhansl CL (1983) The representation of harmonic structure in music: hierarchies of stability as a function of context. *Cognition* 13:63–102.
- Bharucha JJ, Mencl WE (1996) Two issues in auditory cognition: self-organization of octave categories and pitch-invariant pattern recognition. *Psychol Sci* 7:142–149.
- Bharucha JJ, Olney KL (1989) Tonal cognition, artificial intelligence and neural nets. *Contemp Music Rev* 4:341–356.
- Bharucha JJ, Stoeckig K (1987) Priming of chords: spreading activation or overlapping frequency spectra? *Percept Psychophys* 41:519–524.
- Biederman I (1987) Recognition-by-components: a theory of human image understanding. *Psychol Rev* 94:115–147.
- Bigand E (1993) Contributions of music to research on human auditory cognition. In: McAdams S, Bigand E (eds), *Thinking in Sound: The Cognitive Psychology of Human Audition*. Oxford: Clarendon Press, pp. 231–273.
- Bigand E (1997) Perceiving musical stability: the effect of tonal structure, rhythm, and musical expertise. *J Exp Psychol Hum Percept Perform* 23:808–822.
- Bigand E, Pineau M (1996) Context effects on melody recognition: a dynamic interpretation. *Curr Psychol Cogn* 15:121–134.
- Bigand E, Pineau M (1997) Global context effects on musical expectancy. *Percept Psychophys* 59:1098–1107.
- Bigand E, Parncutt R, Lerdahl F (1996) Perception of musical tension in short chord sequences: the influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Percept Psychophys* 58:125–141.
- Bigand E, Perruchet P, Boyer M (1998) Implicit learning of an artificial grammar of musical timbres. *Curr Psychol Cogn* 17:577–600.
- Bigand E, Madurell F, Tillmann B, Pineau M (1999) Effect of global structure and temporal organization on chord processing. *J Exp Psychol Hum Percept Perform* 25: 184–197.
- Bigand E, Tillmann B, Poulin B, D'Adamo DA (2001) The effect of harmonic context on phoneme monitoring in vocal music. *Cognition* 81:B11–B20.
- Bigand E, Poulain B, Tillmann B, D'Adamo D (2003) Cognitive versus sensory components in harmonic priming effects. *J Exp Psychol Hum Percept Perform* 29:159–171.
- Bigand E, Tillmann B, Poulin-Charronnat B, Manderlier D (2005) Repetition priming: is music special? *Q J Exp Psychol*, in press.
- Burns EM (1999) Intervals, scales and tuning In: Deutsch D (ed), *The Psychology of Music*, 2nd ed., San Diego: Academic Press, pp. 215–264.
- Cabeza R, Kingstone A (2001) *Handbook of Functional Neuroimaging in Cognition*. Cambridge, MA: MIT Press.
- Caplan D, Alpert N, Waters G (1999) PET Studies of syntactic processing with auditory sentence presentation. *NeuroImage* 9:343–351.
- Carreras F, Leman M, Lesaffre M (1999) Automatic harmonic description of musical signals using schema-based chord decomposition. *J New Music Res* 28:310–333.
- Castellano MA, Bharucha JJ, Krumhansl CL (1984) Tonal hierarchies in the music of North India. *J Exp Psychol Gen* 113:394–412.

- Chailley J (1951) *Traité historique d'analyse musicale*. Paris: Leduc.
- Chollet SDV (2001) Impact of training on beer flavor perception and description: Are trained and untrained subjects really different? *J Sens Studies* 16:601–618.
- Clark GM (2002) Learning to understand speech with cochlear implant. In Fahle M, Poggio T (eds), *Perceptual Learning*. Cambridge, MA: MIT Press, pp. 147–160.
- Cohen MA, Grossberg S, Wyse LL (1995) A spectral network model of pitch perception. *J Acoust Soc Am* 98:862–879.
- Collins AM, Quillian MR (1969) Retrieval time from semantic memory. *J Verb Learn Verb Behav* 8:241–248.
- Cuddy LL, Thompson WF (1992a) Asymmetry of perceived key movement in chorale sequences: converging evidence from a probe-note analysis. *Psychol Res* 54:51–59.
- Cuddy LL, Thompson WF (1992b) Perceived key movement in four-voice harmony and single voices. *Music Percept* 9:427–438.
- Deutsch, D (1982) (ed) *The Psychology of Music*. New York: Academic Press.
- Deutsch, D (ed) (1999) *The Psychology of Music*. 2nd ed. San Diego: Academic Press.
- DeWitt LA, Samuels AG (1990) The role of knowledge-based expectation in music perception: evidence from musical restoration. *J Exp Psychol Gen* 119:123–144.
- Dowling WJ (1972) Recognition of melodic transformations: inversion, retrograde, and retrograde inversion. *Percept Psychophys* 12:417–421.
- Dowling WJ (1978) Scale and contour: two components of a theory of memory for melodies. *Psychol Rev* 85:341–354.
- Dowling WJ (1986) Context effects on melody recognition: scale-step versus interval representations. *Music Percept* 3:281–296.
- Dowling WJ (1991) Tonal strength and melody recognition after long and short delays. *Percept Psychophys* 50:305–313.
- Dowling WJ, Bartlett JC (1981) The importance of interval information in long-term memory for melodies. *Psychomusicology* 1:30–49.
- Dowling WJ, Harwood DL (1986) *Music Cognition*. Orlando, FL: Academic Press.
- Dowling WJ, Kwak S, Andrews MW (1995) The time course of recognition of novel melodies. *Percept Psychophys* 57:136–149.
- Elman JL, McClelland JL (1984) The interactive activation model of speech perception. In: Lass N (ed), *Language and Speech*. New York: Academic Press, pp. 337–374.
- Embick E, Marantz A, Miyashita Y, O'Neil W, Sakai KL (2000) A syntactic specialization for Broca's area. *Proc NY Acad Sci* 97:6150–6154.
- Fiez JA, Balota DA, Raichle ME, Petersen SE (1999) Effects of lexicality, frequency and spelling-to-sound consistency on the functional anatomy of reading. *Neuron* 24:205–218.
- Fisher GH (1967) Perception of ambiguous stimulus materials. *Percept Psychophys* 2: 421–422.
- Francès R (1958) *La Perception de la Musique*. 2nd ed. Paris: Vrin [1984, *The Perception of Music* (Dowling trans.), Hillsdale, NJ: Earlbaum].
- Friederici AD (1995) The time course of syntactic activation during language processing: a model based on neuropsychological and neurophysiological data. *Brain Lang* 50: 259–281.
- Friederici AD, Meyer M, van Cramon DY (2000) Auditory language comprehension: An event-related fMRI study on the processing of syntactic and language information. *Brain Lang* 74:289–300.
- Fuster JM (2001) The prefrontal cortex—an update: time is of the essence. *Neuron* 30: 319–333.

- Gjerdingen RO (1990) Categorization of musical patterns by self-organizing neuronlike networks. *Music Percept* 8:339–370.
- Griffiths TD, Johnsrude I, Dean JL, Green GGR (1999) A common neural substrate for the analysis of pitch and duration pattern in segmented sound? *NeuroReport* 10:3825–3830.
- Grossberg S (1970) Some networks that can learn, remember and reproduce any number of complicated space-time patterns. *Stud Appli Math* 49:135–166.
- Grossberg S (1976) Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors. *Biol Cybernet* 23:121–134.
- Hebb DO (1949) *The organization of behavior*. New York: John Wiley & Sons.
- Herzog M, Fahle M (2002) Top-down information and models of perceptual learning. In: Fahle M, Poggio T (eds), *Perceptual Learning*. Cambridge, MA: MIT Press, pp. 367–380.
- Howard JH, O'Toole AJ, Parasuraman R, Bennett KB (1984) Pattern-directed attention in uncertain-frequency detection. *Percept Psychophys* 35:256–264.
- Huron D, Parncutt R (1993) An improved model of tonality perception incorporating pitch salience and echoic memory. *Psychomusicology* 12:154–171.
- Janata P (1995) ERP measures assay the degree of expectancy violation of harmonic contexts in music. *J Cogn Neurosci* 7:153–164.
- Janata P, Birk J, Van Horn JD, Leman M, Tillmann B, Bharucha JJ (2002) The cortical topography of tonal structures underlying Western music. *Science* 298:2167–2170.
- Kiehl KA, Laurens KR, Duty TL, Forster BB, Liddle PF (2001) Neural sources involved in auditory target detection and novelty processing: an event-related fMRI study. *Psychophysiology* 38:133–142.
- Koelsch S, Gunter T, Friederici AD (2000) Brain indices of music processing: “non-musicians” are musical. *J Cogn Neurosci* 12:520–541.
- Koelsch S, Gunter TC, v Cramon DY, Zysset S, Lohmann G, Friederici AD (2002) Bach speaks: a cortical “language-network” serves the processing of music. *NeuroImage* 17:956–966.
- Kohonen T (1995) *Self-Organizing Maps*. Berlin: Springer-Verlag.
- Koustaal W, Wagner AD, Rotte M, Maril A, Buckner RL, Schacter DL (2001) Perceptual specificity in visual object priming: functional magnetic resonance imaging evidence for a laterality difference in fusiform cortex. *Neuropsychologia* 39:184–99.
- Krumhansl CL (1990) *Cognitive Foundations of Musical Pitch*. Oxford: Oxford University Press.
- Krumhansl CL, Kessler E (1982) Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychol Rev* 89:334–368.
- Krumhansl CL, Bharucha JJ, Castellano M (1982a) Key distance effects on perceived harmonic structure in music. *Percept Psychophysics* 32:96–108.
- Krumhansl CL, Bharucha JJ, Kessler EJ (1982b) Perceived harmonic structures of chords in three related keys. *J Exp Psychol Hum Percept Perform* 8:24–36.
- Krumhansl CL, Louhivuori J, Toivianinen P, Jarvinen T, Eerola T (1999) Melodic expectation in Finnish spiritual folk hymns: converging evidence of statistical, behavioral and computational analyses. *Music Percept* 17:151–195.
- Kutas M, Hillyard SA (1980) Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biol Psychol* 11:99–116.
- Laden B, Keefe DH (1991) The representation of pitch in a neural net model of chord classification. In: Todd P, Loy G (eds), *Music and Connectionism* Cambridge, MA: MIT Press, pp. 64–83.

- Leman M (1995) *Music and Schema Theory*. Berlin: Springer-Verlag.
- Leman M (2000) An auditory model of the role of short-term memory in probe tone rating. *Music Percept* 17:437–460.
- Leman M, Carreras F (1998) Schema and gestalt: testing the hypothesis of psychoneural isomorphism by computer simulation. In: Leman M (ed). *Music, Gestalt, and Computing*. Berlin: Springer-Verlag, pp. 144–168.
- Leman M, Lessaffre M, Tanghe K (2000) *The IPEM Toolbox Manual*. University of Ghent, IPEM-Dept. of Musicology: IPEM.
- Lerdahl F (1988) Tonal Pitch Space. *Music Percept* 5:315–345.
- Lerdahl F (2001) *Tonal Pitch Space*. New York: Oxford University Press.
- Lerdahl F, Jackendoff R (1983) *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.
- Linden DEJ, Prvulovic D, Formisano E, Vollinger M, Zanella FE, Goebel R, Dierks T (1999) The functional neuroanatomy of target detection: an fMRI study of visual and auditory oddball tasks. *Cereb Cortex* 9:815–823.
- Maess B, Koelsch S, Gunter T, Friederici AD (2001) ‘Musical syntax’ is processed in the Broca’s area: an MEG-study. *Nat Neurosci* 4:540–545.
- Marr D (1982) *Vision: a computational investigation into the human representation and processing of visual information*. San Francisco: Freeman.
- McAdams S, Bigand E (1993) *Thinking in Sound*. Oxford: Clarendon Press.
- McClelland JL, Elman JL (1986) The TRACE model of speech perception. *Cogn Psychol* 18:1–86.
- McClelland JL, Rumelhart DE (1981) An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychol Rev* 86:287–330.
- McClelland JL, Rumelhart DE (1986) *Parallel distributed processing: Exploration in the Microstructure of Cognition, Vol. 2*. Cambridge, MA: MIT Press.
- Moore DR (2002) Auditory development and the role of experience. *Br Med Bull* 63: 171–181.
- Morrot G, Brochet F, Dubourdieu D (2001) The colors of odors. *Brain Lang* 78:309–320.
- Müller R-A, Kleinhans N, Courchesne E (2001) Broca’s area and the discrimination of frequency transitions: a functional MRI study. *Brain Lang* 76:70–76.
- Page MA (1994) Modeling the perception of musical sequences with self-organizing neural networks. *Connect Sci* 6:223–246.
- Palmer C, Krumhansl CL (1987a) Independent temporal and pitch structures in determination of musical phrases. *J Exp Psychol Hum Percept Perform* 13:116–126.
- Palmer C, Krumhansl CL (1987b) Pitch and temporal contributions to musical phrase perception: effects of harmony, performance timing, and familiarity. *Percept Psychophys* 41:505–518.
- Parks TN, Rubel AW, Popper AN, Fay RR (eds) (2004) *Springer Handbook of Auditory Research, Vol. 23: Plasticity of the Auditory System*. New York: Springer-Verlag.
- Parncutt R (1988) Revision of Terhardt’s psychoacoustical model of the roots of a musical chord. *Music Percept* 6:65–94.
- Parncutt R (1989) *Harmony: a psychoacoustical approach*. Berlin: Springer-Verlag.
- Patel AD, Gibson E, Ratner J, Besson M, Holcomb PJ (1998) Processing syntactic relations in language and music: an event-related potential study. *J Cogn Neurosci* 10: 717–733.
- Perruchet P, Pacteau C (1990) Synthetic grammar learning: implicit rule abstraction or explicit fragmentary knowledge? *J Exp Psychol Gen* 119:264–275.

- Perruchet P, Vinter A (2002) The self-organizing consciousness. *Behav Brain Sci* 25: 297–330.
- Perruchet P, Vinter A, Gallego J (1997) Implicit learning shapes new conscious percepts and representations. *Psychon Bull Rev* 4:43–48.
- Philibert B, Collet L, Vesson J, Veuillet E (2002) Intensity-related performances are modified by long-term hearing aid use: a functional plasticity? *Hear Res* 165:142–151.
- Poldrack RA, Wagner AD, Prull MW, Desmond JE, Glover GH, Gabrieli JDE (1999) Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *NeuroImage* 10:15–35.
- Pugh KR, Shaywitz BA, Fulbright RK, Byrd D, Skudlarski P, Katz L, Constable RT, Fletcher J, Lacadie C, Marchione K, Gore JC (1996) Auditory selective attention: an fMRI investigation. *NeuroImage* 4:159–173.
- Rameau J-P (1721) *Treatise on Harmony* (Gosset P, Trans.) (1971 ed.). New York: Dover.
- Reber AS (1967) Implicit learning of artificial grammars. *J Verb Learn Verb Behav* 6: 855–863.
- Reber AS (1989) Implicit learning and tacit knowledge. *J Exp Psychol Gen* 118:219–235.
- Reber AS (1992) The cognitive unconscious: an evolutionary perspective. *Consc Cogn* 1:93–133.
- Reber AS, Walkenfeld F, Hernstadt R (1991) Implicit and explicit learning: individual differences and IQ. *J Exp Psych Learn Mem Cogn* 17:888–896.
- Regnault P, Bigand E, Besson M (2001) Event-related brain potentials show top-down and bottom-up modulations of musical expectations. *J Cogn Neurosci* 13:241–255.
- Rosch E (1975) Cognitive reference points. *Cogn Psychol* 7:532–547.
- Rosch E (1979) On the internal structure of perceptual and semantic categories. In: Moore TE (ed), *Cognitive Development and the Acquisition of Language*. New York: Academic Press.
- Rosen C (1971) *Le style classique, Haydn, Mozart, Beethoven* (M. Vignal, trans). Paris: Gallimard.
- Rubel EW, Popper AN, Fay RR (eds) (1997) *Springer Handbook of Auditory Research, Vol. 9: Development of the Auditory System*. New York: Springer-Verlag.
- Rumelhart DE, McClelland JL (1982) An interactive activation model of context effects in letter perception. Part 2. *Psychol Rev* 89:60–94.
- Rumelhart DE, Zipser D (1985) Feature discovery by competitive learning. *Cogn Sci* 9: 75–112.
- Saffran J, Aslin R, Newport E (1996) Statistical learning by 8-month-old infants. *Science* 274:1926–1928.
- Saffran JR, Johnson EK, Aslin RN, Newport EL (1999) Statistical learning of tone sequences by human infants and adults. *Cognition* 70:27–52.
- Saffran JR, Newport EL, Aslin RN, Tunick RA, Barrueco S (1997) Incidental language learning. *Psychol Sci* 8:101–105.
- Sano H, Jenkins BK (1991) A neural network model for pitch perception. In Todd N, Loy G (eds), *Music and Connectionism*. Cambridge, MA: MIT Press, pp. 42–49.
- Sasaki T (1980) Sound restoration and temporal localization of noise in speech and music sounds. *Tohoku Psychol Folia* 39:70–88.
- Schenker H (1935) *Der Freie Satz. Neue musikalische Theorien und Phantasien* (N Meeüs, Trans.) Liège: Margada.
- Shepard RN (1964) Circularity in judgments of relative pitch. *J Acoust Soc Am* 36: 2346–2353.

- Taylor I, Greenhough M (1994) Modeling pitch perception with adaptive resonance theory artificial neural networks. *Connect Sci* 6:135–154.
- Tekman HG, Bharucha JJ (1992) Time course of chord priming. *Percept Psychophys* 51:33–39.
- Tekman HG, Bharucha JJ (1998) Implicit knowledge versus psychoacoustic similarity in priming of chords. *J Exp Psychol Hum Percept Perform* 24:252–260.
- Thompson WF, Cuddy LL (1989) Sensitivity to key change in chorale sequences: a comparison of single voices and four-voice harmony. *Music Percept* 7:151–168.
- Tillmann B, Bharucha JJ (2002) Harmonic context effect on temporal asynchrony detection. *Percept Psychophys* 64:640–649.
- Tillmann B, Bigand E (2001) Global relatedness effect in normal and scrambled chord sequences. *J Exp Psychol Hum Percept Perform* 27:1185–1196.
- Tillmann B, Bigand E, Pineau M (1998) Effects of global and local contexts on harmonic expectancy. *Music Percept* 16:99–117.
- Tillmann B, Bharucha JJ, Bigand E (2000) Implicit learning of tonality: a self-organizing approach. *Psychol Rev* 107:885–913.
- Tillmann B, Bharucha JJ, Bigand E (2001) Implicit learning of regularities in Western tonal music by self-organization. In: French R, Sourné, J (eds), *Proceedings of the Sixth Neural Computation and Psychology Workshop: Evolution, Learning, and Development. Perspectives in Neural Computing*. London: Springer-Verlag, pp. 175–184.
- Tillmann B, Janata P, Bharucha JJ (2003) Inferior frontal cortex activation in musical priming. *Cogn Brain Res* 16:145–161.
- Tillmann B, Bigand E, Escoffier N, Lalitte P (2004) Influence of harmonic context on musical timbre processing. Manuscript submitted for publication.
- von Helmholtz HL (1885/1954) *On the sensations of tone as a physiological basis for the theory of music* (A.J. Ellis, Trans.) London: Longmans, Green.
- Wagner AD, Koustaal W, Maril A, Schacter DL, Buckner RL (2000) Task-specific repetition priming in left inferior prefrontal cortex. *Cereb Cortex* 10:1176–1184.
- Wagner AD, Paré-Blagoev EJ, Clark J, Poldrack RA (2001) Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. *Neuron* 31:329–338.
- Warren RM (1970) Perceptual restoration of missing speech sounds. *Science* 167:392–393.
- Warren RM (1999) *Auditory perception: a new analysis and synthesis*. Cambridge, UK: Cambridge University Press.
- Warren RM, Sherman GL (1974) Phonemic restoration based on subsequent context. *Percept Psychophys* 16:150–156.
- West WC, Dale AM, Greve D, Kuperberg G, Waters G, Caplan D (2000) Cortical activation during a semantic priming lexical decision task as revealed by event-related fMRI. Paper presented at the Human Brain Mapping Meeting. Poster #360, *Neuro-Image* 360.
- Zatorre RJ, Evans AC, Meyer E, Gjedde A (1992) Lateralization of phonetic and pitch processing in speech perception. *Science* 256:846–849.
- Zatorre RJ, Evans AC, Meyer E (1994) Neural mechanisms underlying melodic perception and memory for pitch. *J Neurosci* 14:1908–1919.

# Index

- Absolute pitch, 213
  - in starling, 83
- Absolute thresholds, and frequency difference limen, 236
- Acoustic environments, nonhuman animals, 91
- Active cochlear mechanisms, pitch theory, 187
- Aer internus, pitch theory, 191
- All-order interspike interval histogram, autocorrelation, 196
- AM (amplitude modulation), and frequency discrimination, 235–236
- Amplitude modulated noise, pitch, 33–34
- Analytic listening, partial pitch, 16ff, 184
  - pitch in nonhuman animals, 69ff
- Anterior superior temporal lobe, pitch, 158
- Attention, sequential streaming, 299
- Auditory cortex, auditory perception, 129ff
  - Callithrix*, 132–133
  - descending projections, 133–134
  - gerbil, 132
  - Macaca*, 131
  - missing fundamental pitch, 68
  - pitch perception, 129, 131–132, 136–137
  - representation of frequency, 130–131
  - thalamic projections, 133
  - tonotopy, 130–131, 153ff
- Auditory filter, bandwidth, 20–21
  - guinea pig, 105
  - humans, 105
- Auditory filter shapes, and frequency difference limens, 236–237
- Auditory grouping, 278ff
- Auditory image model (AIM), 152, 198–199
- Auditory nerve, cat, 100
  - fiber groups, 100–101
  - phase locking, 102–104
  - pitch, 100ff
  - representation of formant peaks, 109
  - representation of fundamental frequency, 107
  - representation of vowels, 110
  - response to harmonic complex sounds, 106–107
  - response to narrow-band stimuli, 35–36
  - spontaneous rate, 100–101
- Auditory streaming, 296–298, 300
  - pitch perception in nonhuman animals, 87ff
- Autocorrelation, definition of, 193
  - pitch perception, 108, 136, 282, 283
  - and pitch theory, 193ff
- Autocorrelation models, click trains, 110–112
  - nerve activity, 35
  - pitch, 30–31, 35, 172
- Autocorrelation vs. pattern matching, pitch models, 172
- Average localized synchronized rate (ALSR), pitch theory, 188
- Average magnitude difference function (AMDF), pitch theory, 197



- Basilar membrane**, excitation by complex tone, 21–22  
 fundamental frequency, 105  
 pitch, 11  
 vibration patterns, 105
- Beethoven's 9th symphony, 316
- Bell strike note, pitch theory, 192–193
- Best modulation frequency, inferior colliculus, 127–129
- Binaural coherence edge pitch, 39
- Binaural effects, pitch theory, 214–215
- Binaural processing, 36
- BOLD** (blood oxygenation level dependence), and dendritic input activity, 148  
 technique, 148ff
- BOLD SNR**, brainstem and cortex, 149
- Bottom-up processes, 306, 307, 308ff
- Brain, music, 332–333
- Brain function, harmonic relationships, 333
- Brain imaging, see also fMRI, PET  
 individual differences, 163  
 pitch, 147ff  
 response to clicks, 161
- Broca's area, music, 332–333
- Budgerigar, see also Parakeet  
 rippled noise perception, 72ff
- Bullfrog, see also *Rana catesbeiana*  
 harmonic complex starting phase, 78–79  
 responses to mistuned harmonics, 69
- Callithrix jacchus jacchus* (marmoset),  
 primary auditory cortex, 132–133
- Cancellation model of pitch, 197
- Carassius auratus*, see Goldfish
- Cardiac triggering, brainstem imaging, 149–150
- Carp, see *Cyprinus carpio*
- Cat, auditory nerve, 100  
 interspike intervals, 108  
 missing fundamental and cortex, 68  
 missing fundamental perception, 66ff  
 modulation transfer function in inferior colliculus, 128  
 olivocochlear system, 109  
 receptive fields in auditory cortex, 132–133  
 spike discharge patterns along cochlea, 101–102
- Cebus apella* (cebus monkey), frequency contour perception, 83–84
- Cell types, cochlear nucleus, 113ff  
 inferior colliculus, 127
- Cepstral analysis, 107
- Chinchilla, dominance region for pitch, 76–77  
 frequency difference limens, 63  
 listening strategy, 75–76  
 modulation rate discrimination, 64  
 noise coloration discrimination, 74  
 pitch strength effects, 80  
 rippled noise generalization, 72, 74–75
- Chopper units, cochlear nucleus, 116ff  
 concurrent vowels, 118  
 gerbil, 120  
 periodicity coding, 118–119  
 and pitch theory, 216  
 representation of formants, 116, 135–136  
 rippled noise, 125  
 selective listening, 116–117
- Chords, music, 325ff  
 processing, 325ff
- Chroma, octaves, 177, 213  
 pitch, 176–177
- Chromatic scale, 315
- Classical conditioning, animal perception, 57ff
- Click train(s), representation of F0 of, 110–112  
 MEG, 162
- CNS, processing of pitch, 5
- Cochlea, mapping of pitch, 3  
 nonlinearities, 14, 106  
 spike discharge patterns, 101–102
- Cochlear implants, and pitch perception, 29, 234ff, 256ff
- Cochlear length, pitch perception, 89
- Cochlear map, cat, 90  
 chinchilla, 90  
 guinea pig, 90  
 human, 90  
 monkey, 90
- Cochlear nucleus, 112ff, see also Chopper units, DCN, Onset units, Primary-like units, VCN

- cell types, 113ff
- chopper units, 116ff
- coding of periodicity, 118–119, 121–122
- concurrent vowels, 118
- inhibition in, 121
- octopus cells, 121–122
- onset units, 119ff
- periodicity coding, 118–119
- primary-like units, 114–116
- range of periodicities, 119
- regions, 112
- representation of fundamental frequency, 114
- response in gerbil, 120
- response to iterated rippled noise, 123, 125–126
- selective listening, 116–117
- tonotopy, 112
- Cognition, Western music, 306ff
- Columbia livia* (pigeon), pure tone frequency generalization, 61
- Combination tones, and pitch theory, 14, 202ff
- Complex sounds, representation of F0 of, 105ff, 109, 279–280
- Computer models, pitch theory, 216–217
- Concurrent vowels, representation in cochlear nucleus, 118
- Conditioning methods, animal perception, 57ff
- Context effects, neurophysiological basis, 330ff
  - on pitch perception, 287–288, 311ff
- Contextual distance, 318
- Continuous interleaved sampling (CIS), 257
- Cortical neurons, spontaneous rate, 130
- Cowbird, see *Molothrus* sp.
- Cross correlation, pitch theory, 184
- Cubic difference tone, pitch theory, 202–203
- Cyprinus carpio* (carp), music discrimination, 80ff
- Cytoarchitecture, inferior colliculus, 127
- Dante's Inferno**, 111
- DCN, representation of pitch, 122–123
- Dead region, of low frequency, 241
  - and pitch perception, 244
- Definition, pitch, 1–2
  - pure tone, 8
  - resolvability, 20
- Descending projections, from auditory cortex, 133–134
  - Pteronotus parnelli parnelli*, 134
- Dichotic pitch, 38–40
- Dichotic presentation, pitch, 27–29
- Difference limens, measurement, 7–8
- Difference tone, pitch theory, 202
- Diplacusis, and hearing loss, 246
- Discrimination, effect of duration, 42–44
  - F0, 18–20, 24–25
- Discrimination learning, 58ff
- Dissonance, music, 312
- Distortion product otoacoustic emissions, 14
- Dolphin, see also *Tursiops truncatus*
  - dominance region for pitch, 76
- Dominance region for pitch, nonhuman animals, 76ff
  - pitch perception, 14–16
- Dominant component model, 199
  - central filterbank, 188
- Dorsal cochlear nucleus, see Cochlear nucleus, DCN
- Dual mechanism hypotheses, pitch theory, 209–210
- Duration, effect on discrimination, 42–44
  - integration mechanism, 45–47
- Dynamic pitch, 206–207
- EEG** (electroencephalography), Laplacian models, 159
  - latencies of response, 161
  - pitch, 147ff
- Eighth nerve, see Auditory nerve
- Electromagnetic techniques (EEG and MEG), 159ff
- Encoding pitch, 37–38, 99ff
  - F0 of complex sounds, 105ff
  - onset units of cochlear nucleus, 121
  - representation of F0 in presence of competing sounds, 109–110
- Envelope, transposed fine structure, 35–36
- Envelope prominence, chinchilla perception, 78

- Envelope-weighted average instantaneous frequency (EWAIF), 207
- Equalization and cancellation model, 214
- Excitation pattern, frequency discrimination, 235
- Excitation pattern edge, pitch theory, 188
- Existence regions, pitch, 26ff
- F0**, see also Fundamental frequency, Fundamental frequency difference limen  
 effect of duration, 42–44  
 effects of training, 39  
 as function of F0, 23–24  
 harmonics, 44–45  
 spectral response, 37–38  
 vs. frequency discrimination limen, 19, 23
- F0 differences, changes in speech, 295  
 discrimination, 18–20, 24–25  
 frequency modulation effects, 295–296  
 global vs. local comparison, 18–20  
 harmonics, 14–16  
 improvement in intelligibility, 292–294  
 localization, 294–295  
 necessity for pitch perception, 13–14  
 pitch, 14–16  
 relationship to F0DL, 23–24  
 relationship to music, 312–313  
 resolvability of sounds, 23–24
- FDL, see Frequency difference limen
- First effect of pitch shift, 185
- First-order interspike interval histogram, 196
- FM (frequency modulation) and frequency discrimination, 235–236
- FMDL (frequency modulation detection limen), and hearing impairment, 237ff
- fMRI (functional magnetic resonance imaging), see also Brain imaging  
 music, 330, 333  
 pitch, 147ff
- fMRI BOLD activation, human brain image, 150
- Formant, definition of, 175–176  
 representation in cochlear nucleus neurons, 135–136  
 representation of peaks in auditory nerve, 109
- Formant frequency, 107
- Formant peaks, representation in chopper units, 116
- Fourcin pitch, 39
- Fourier analysis, pitch theory, 169
- Fourier transformer, cochlea, 181
- Frequency, effect on frequency difference limen, 10
- Frequency contour perception, nonhuman animals, 82ff
- Frequency difference limen, 7–8  
 and absolute thresholds, 236  
 and auditory filter shapes, 236–237  
 effect of sound level, 10–11  
 frequency dependence, 10  
 and psychophysical tuning curves, 236  
 pure tones, 10–11  
 variability, 236  
 vs. F0DL, 19, 23
- Frequency discrimination, animals, 60ff  
 effect of duration, 42–44  
 physiological basis, 104  
 and pitch perception, 235ff  
 and pitch theory, 190–191  
 spontaneous rate, 104
- Frequency modulation detection limen (FMDL), F0, 295–296  
 and hearing impairment, 237ff
- Frequency representation, auditory cortex, 130–131
- Functional brain imaging, 147ff
- Fundamental frequency, see also F0  
 cat, 106–107  
 effect of masking, 110  
 interspike interval representation, 135  
 location on basilar membrane, 105  
 modulation, 41  
 primary-like units, 114  
 rate-place theory, 105  
 representation in auditory nerve, 107  
 representation in auditory cortex, 131–132
- Fundamental frequency difference limen, 7–8, see also F0DL
- Fundamental frequency discrimination, and component phases, 249ff

- and hearing impairment, 249ff
- nonhuman animals, 63ff
- Fundamental frequency profiles, over time, 158
- Gabor, time–frequency tradeoff, 204
- Generalization paradigms, pitch perception in animals, 59–60
- Gerbil, missing fundamental and cortex, 68
  - response of auditory cortex, 132
  - response of cochlear nucleus, 120
- Global pitch, 16ff
- Goldfish, analytic listening, 71
  - missing fundamental pitch, 65–66
  - modulation rate discrimination, 64
  - octave generalization, 82
  - pulse repetition rate perception, 66
  - pure tone frequency generalization, 61, 65–66
  - rippled noise perception, 72ff
  - SAM tone perception, 65
  - timbre perception, 66
- Goldstein, pitch model, 183
- Go/No Go paradigm, 58
- Grammar, musical, 312
- Group frequency, 204
- Grouping, pitch perception, 279ff, 291ff
- Guinea pig, auditory filter, 105
  - maximum spike rate, 191
  - phase locking, 103
  - response of VCN to iterated rippled noise, 126
- Harmonic complex**, definition of, 175
- Harmonic component phase, and pitch, 249
- Harmonic expectancies, 324
- Harmonic number, F0 discrimination, 24–25
  - resolved vs. unresolved harmonics, 22, 23
- Harmonic priming, 325
- Harmonic relationships, brain function, 333
- Harmonic sieve, 184, 211, 280–282
- Harmonic template, origins, 185–186
- Harmonicity, perception of timbre, 291ff
  - pitch perception, 280ff
- Harmonics, see also Resolved harmonics, Unresolved harmonics
  - contributions to periodicity pitch, 28
  - dominant, 14–16
  - effects on F0DL, 44–45
  - F0, 14–16
  - pitch, 14ff, 32–33
  - resolved, 20ff
  - response of auditory nerve, 106–107
  - unresolved, 20ff
  - use by humans in pitch perception, 280
- Harmonics resolved, and pitch salience, 156
- Harmony, and pitch, 212ff
  - Western tonal music, 322–324
- Hearing, binaural processing, 36
- Hearing impairment, and pitch perception, 29, 234ff
- Hell, fifth level, 111
- Helmholtz, auditory theory, 181
  - views of pitch, 13
- Hemodynamic imaging, pitch, 147ff
- Hemodynamic responses, speed, 148
- Hemodynamic techniques (PET and fMRI), 147ff
- Heschl's gyrus, MEG imaging, 160
  - primary auditory cortex, 153
- Huggins pitch, 39
- Human, auditory filter, 105
  - pitch and localization cues, 289–291
  - pitch perception, 278ff
  - sound localization, 289–291
  - use of harmonics in pitch perception, 280
- Human auditory cortex, homolog to area R, 157–158
- Hyla cinerea* (green treefrog), detection of harmonic sounds, 69–70
- Imaging**, brain, 147ff
- Implicit learning, 334ff
  - neural net modeling, 338ff
  - pitch regularities, 334ff
- Implied polyphony, 297
- Individual differences, pitch brain imaging, 158

- Inferior colliculus, 125ff  
 best modulation frequency, 127–129  
 cell types, 127  
 cytoarchitecture, 127  
 modulation transfer function, 127–128  
 periodicity tuning, 127–129  
 rate-level functions, 127  
 tonotopic organization, 127–128
- Inhibition, onset units of cochlear nucleus, 121
- Inner hair cells missing, pitch perception, 235ff
- Instrumental avoidance conditioning, animal perception, 58
- Integration, nonsimultaneous harmonics, 44–45
- Integration mechanism, 45–47  
 duration, 45–47  
 multiple looks, 45–47  
 resetting, 46–47  
 unresolved harmonics, 46–47
- Intensity-weighted average instantaneous frequency (IWAIF), 207
- Interaural time difference, effect on F0  
 identification, 294–295  
 human, 289–291
- Internal noise, 45
- Interspike intervals, cat, 108  
 pitch representation, 107–108  
 and pitch theory, 196–197  
 representation of fundamental frequency, 135
- ITD, see Interaural time difference
- Iterated rippled noise, brain imaging, 151  
 pitch, 34–35  
 response in cochlear nucleus, 123ff  
 response in VCN, 126
- Japanese macaque, auditory streaming, 87
- Keys**, music, 316–317
- Knowledge-based processes, 306, 307, 308ff
- Knowledge-driven expectancy, pitch structure, 324ff
- Late positive component, see P300 potential
- Lateral inhibitory network (LIN), and pitch theory, 188
- Lateralization, of pitch mechanisms, 158
- Learning, pitch theory, 185–186
- Level, effects on pitch, 12–13
- Licklider, autocorrelation theory, 193–194  
 triplex pitch theory, 214
- Linear time-invariant system, 174
- Linearity, and principle of superposition, 179
- Listening, analytic, 16ff  
 synthetic, 16ff
- Listening experience, nonhuman animals, 91
- Localization, effects of differences in F0, 294–295
- Localization cues, pitch perception, 288ff
- Macaca* sp. (macaque monkey), A1 and R in cortex, 155  
 consonance perception, 81–82  
 missing fundamental and cortex, 68  
 modulation frequency discrimination, 64–65  
 modulation rate discrimination, 64  
 response of auditory cortex, 131
- Macaque monkey, see *Macaca* sp.
- Mapping, and pitch theory, 217
- Marmoset, see *Callithrix jacchus jacchus*
- Masking, effect on fundamental frequency, 110
- Matched filters, 201
- Maximum integration time, 206
- Mean rate model, pitch, 31
- MEG (magnetoencephalography), latencies of response, 161  
 nonunique solutions to the inverse problem, 158  
 pitch, 147ff  
 and tonotopy, 161–162
- Mel scale, 9
- Mel spectrum, pitch pattern matching, 184
- Melodies, pitch, 39–40
- Melody, and pitch, 212ff  
 vs. random pitch, 158
- Melody discrimination, monkey, 85ff
- Melody perception, 320–321

- Melopsitacus undulatus* (parakeet), pitch scale, 61–62
- Memorization, pitch structure, 311ff, 319
- Meriones unguiculatus* (gerbil), SAM perception, 63–64
- Method of best beats, 203
- Method of constant stimuli, 8
- Methods, difference limens, 7–8  
method of constant stimuli, 8
- Middle latency response, and Nam MEG, 161
- Mimus* sp. (mockingbird), pitch perception, 83
- Missing fundamental, 13–14
- Missing fundamental experiments, 149
- Missing fundamental perception, brain mechanisms, 163  
nonhuman animals, 65ff
- Missing fundamental pitch, auditory cortex of cats, 68  
and auditory nonlinearities, 69  
cat, 66ff  
monkey, 67  
pattern matching models, 183ff  
zebra finch, 71
- Missing fundamental stimuli, MEG, 162
- Mistuned harmonic, 283–284  
detection thresholds in birds vs. humans, 70–71  
pitch in nonhuman animals, 69ff
- Model, definition of, 172–173
- Model for pitch, stretched string, 182
- Modeling, Western pitch structures, 338ff
- Modulation filterbank, pitch theory, 188, 199
- Modulation frequency discrimination, nonhuman animals, 63ff
- Modulation rates, pitch, 40–42
- Modulation sensitivity, temporal integration, 40–42
- Modulation transfer function, inferior colliculus, 127–128
- Molothrus* sp. (cowbird), pitch perception, 83
- Monkey, see also Macaque monkey  
frequency contour perception, 83–84  
octave generalization, 86  
SAM vs. QFM, 78
- Müller, specific nerve energies, 183
- Multidimensional scaling, rate and place, 261
- Multiple looks strategy, 45–47, 205
- Multiple pitches, 210ff
- Music, brain, 332–333  
chords, 314–315, 322, 325ff  
chromatic scale, 315  
context effects, 330ff  
definitions of pitch, 1  
dissonance, 312  
fMRI, 333  
homophonic, 284  
keys, 316–317  
neurophysiology, 330ff  
notes, 314ff, 322  
P300 potential, 330–332  
perception of, 311ff  
pitch perception, 1–2, 312ff  
polyphonic, 284  
prefrontal cortex, 332–333  
psychoacoustic constraints, 311–312
- Music discrimination, nonhuman animals, 80ff
- Musical grammar, 312
- Musical melody, relationship to pitch, 9
- Musical pitch, cognitive foundation, 306ff, 319  
and place of excitation cues, 265  
and temporal cues, 262
- Musical priming, 334
- Mustached bat, see *Pteronotus parnelli parnelli*
- Narrowed autocorrelation function (NAC) model, 199
- Neural activity, autocorrelation function model, 35
- Neural code evaluation, nonhuman animals, 92
- Neural net, 338ff  
self-organizing maps (SOMs), 338ff
- Neural net modeling, implicit learning, 338ff
- Neural representation of pitch, 100
- Neurophysiology, musical perception, 330ff  
pitch, 99ff
- Noise, periodicity pitch, 28–29, 33–34

- Noise coloration discrimination, chinchilla, 74
- Nonhuman pitch perception, 56ff
- Nonlinearities, cochlea, 14, 106
- Nonsimultaneous harmonics, integration, 44–45
- Notes, musical, 314ff
- Novelty, response of neurons in auditory cortex, 130
  
- Octave equivalence**, 213
- Octave generalization, nonhuman animals, 82ff, 86
- Octopus cells, 121–122
- Ohm, perception of pitch, 13, 180
- Old plus new heuristic, 286
- Olivocochlear system, 109
- Onset synchrony, influence on pitch perception, 284ff
- Onset time, influence on pitch perception, 284ff
- Onset units, cochlear nucleus, 119ff
- Operant conditioning, animal perception, 58
- Otoacoustic emissions, distortion product, 14
  
- P300 potential**, music, 330–332
- Parakeet, detection of harmonic sounds, 70–71
  - modulation rate discrimination, 64
  - pitch perception, 83
  - starting phase discrimination, 79–80
- Pattern matching models, pitch perception, 170–171
  - pitch theory, 185ff
- Pattern recognition models, residue pitch, 105
- Perception, effects of context, 311ff
  - melody, 320–321
  - models, 306–307
  - residue pitch, 105
  - visual, 308ff
  - Western tonal music, 306ff
- Perception of musical intervals, and hearing loss, 253
- Period histogram, pitch theory, 184
- Periodicity, auditory cortex, 132
  - coding in cochlear nucleus, 121–122
  - range in cochlear nucleus, 119
- Periodicity coding, cochlear nucleus, 118–119
- Periodicity discrimination, nonhuman animals, 60ff
- Periodicity perception, phase effects, 76ff
- Periodicity pitch, 192
  - contribution of harmonics, 28
  - contributions of noise, 28–29
  - number of components needed, 27–29
  - vs. spectral pitch, 176
- Periodicity tuning, inferior colliculus, 127–129
- Peripheral activity pattern, Wightman's model, 183–184
- PET (positron emission tomography), music physiology, 330
  - pitch, 147ff
- Phase, unresolved harmonics, 32–33
- Phase effects, periodicity perception, 76ff
- Phase frequency, 204
- Phase locking, 102–104
  - guinea pig, 103
  - pure tones, 12
  - reduced precision of, and pitch perception, 234ff
- Phase sensitivity, and pitch theory, 192, 195–196
- Phase-opponency model, 102
- Phonemic restoration, 310
- Physiological bandwidth, relationship to psychophysical bandwidth, 106
- Physiological models of pitch, 215–216
- Pigeon, music discrimination, 80ff
  - pitch perception, 83
  - pure tone frequency generalization, 61
- Pitch, amplitude modulated noise, 33–34
  - and auditory grouping, 278ff
  - autocorrelation function models, 30–31, 35
  - basilar membrane, 11
  - binaural edge, 39
  - binaural processing, 36
  - coding in ear, 11ff
  - cognition of musical, 306ff, 319
  - complex sound, 13ff, 279–280

- definition of, 1–2, 176–177
- dichotic presentation, 27–29, 38–40
- effect of combination tones, 14
- encoding, 37–38, 99ff
- encoding in onset units of cochlear nucleus, 121
- existence regions, 26ff
- extracting from mixtures of sound, 284ff
- F0, 114–116
- global, 16ff
- grouping sounds, 291ff
- harmonic effects, 32–33
- in hearing impaired, 29
- Helmholtz, 13
- Huggins, 39
- importance of temporal context, 317–319
- individual harmonics, 17–18
- iterated rippled noise, 34–35
- level dependence, 12–13
- lower limits, 26–27
- mapping in cochlea, 3
- mean rate model, 31
- mechanism, 27–28, 36–38
- modulation rate, 40–42
- and music, 1–2
- in music, 312ff
- neurophysiology, 99ff
- Ohm, 13
- perception, 306ff
- perception in music, 312
- place vs. temporal coding, 11ff
- processing in the CNS, 5
- psychophysics, 7ff
- pure tones, 9–10
- rate-place code, 101–102
- relationship to musical melody, 9
- relationship to “place,” 101
- representation by rate code, 109
- representation in DCN, 122–123
- representation of F0 in presence of competing sounds, 109–110
- representation of F0 of complex sounds, 105ff
- sequential grouping of speech, 299–300
- sequential presentation, 27–29
- temporal integration, 40ff
- two-component complexes, 27–28
- unresolved harmonics, 4, 29ff, 287
- upper limits, 26
- Pitch anomalies, 246ff
  - and music perception, 256
- Pitch center, in secondary auditory cortex, 155ff
- Pitch chroma, 312–314
- Pitch classes, Western music, 313–314
- Pitch coding, auditory cortex, 136–137
- Pitch contour tracking, MEG, 164
- Pitch of complex sounds, auditory cortex, 129ff
  - and hearing loss, 248ff
- Pitch of pure tones, cochlear impairment, 241ff
- Pitch pattern perception, brain imaging, 163–164
- Pitch perception alterations, cochlear implants, 266
- Pitch profiles, over time, 158
- Pitch regularities, implicit learning, 334ff
- Pitch representation, interspike intervals, 107–108
- Pitch salience, and brain imaging, 156ff
- Pitch scale, parakeet vs. human, 61–62
- Pitch shift, and cochlear hearing loss, 247
  - pitch theory, 191–192
- Pitch strength, human vs. nonhuman perceptions, 80
- Pitch structure, dependence on knowledge-based expectations, 311
  - effect of context, 306ff
  - knowledge-driven expectancy, 324ff
  - memorization, 311ff, 319
- Pitch temporal models, and hemodynamic imaging, 148
- Pitch theories, spectral, 169ff
  - waveform, 169ff
- Pitch theory, and color vision, 218
- Pitch-matching, and cochlear impairment, 242ff
- Place, pitch, 11ff
- Place cues, cochlear implants, 260ff
- Place of excitation pitch, cochlear implants, 264–265
- Place theory, early versions, 178ff
  - of Helmholtz, 181–182
- Polyphonic music, 284



- Posterior superior temporal lobe, pitch, 158
- Prefrontal cortex, tonal material, 332–333
- Primary-like units, cochlear nucleus, 114–116  
     representation of formants, 135–136  
     rippled noise, 125
- Priming effects, 324–325
- Principal cells, inferior colliculus, 127
- Processing, musical chords, 325ff
- Processing of pitch structure, knowledge-driven expectancy, 324ff
- Profile analysis, zebra finch, 71
- Psychophysical bandwidth, relationship to physiology bandwidth, 106
- Psychophysical tuning curve, and dead region, 241  
     and frequency difference limens, 236
- Psychophysics, methodology, 7–8  
     pitch, 7ff
- Pteronotus parnelli parnelli* (mustached bat), descending projections of auditory cortex, 134
- Pulse rate discrimination, and cochlear implants, 261–262
- Pulse repetition rate, goldfish, 66
- Pure tone, definition, 8  
     frequency difference limen, 10–11  
     phase locking, 12  
     pitch, 9–10
- Pure tone frequency discrimination, pitch perception, 235ff
- Pure tone frequency generalization, various vertebrate species, 61
- Pure tone perception, animals, 60ff
- Pure tone pitch, cochlear impairment, 241ff
- Ramp and damp perception, 61**
- Rana catesbeiana* (bullfrog), see also Bullfrog  
     synthetic mating call pitch, 63
- Rat, frequency contour perception, 83–84  
     music discrimination, 80ff  
     octave generalization, 82  
     pure tone frequency generalization, 61
- Rate code, representation of pitch, 109
- Rate-level functions, inferior colliculus, 127
- Rate-place code, 101–102  
     effect of background noise, 109
- Rate-place information, frequency discrimination, 104
- Rate-place theory, fundamental frequency, 105
- Receptive fields, auditory cortex, 132–133
- Repetition priming, 324–325
- Residue pitch, 105, 192
- Resolvability, definition, 20  
     F0 discrimination, 24–25
- Resolved harmonics, 20ff  
     pitch mechanism, 36–38
- Resolved vs. unresolved partials, 20ff, 207–208
- Resonance theory, place theory, 178–179
- Rhesus monkey, melody discrimination, 85ff
- Rippled noise, 34–35, 123  
     chopper units, 125  
     primary-like units, 125  
     temporal representation in cochlear nucleus, 125
- Rippled noise detection/discrimination, budgerigar, 72ff  
     chinchilla, 72ff  
     goldfish, 72ff
- Rippled noise generalization, chinchilla, 74–75
- Rippled noise perception, nonhuman animals, 72ff
- Rutherford, telephone theory, 191
- SAM noise pitch, 64**
- SAM vs. QFM, monkeys, 78
- Schroeder-phase stimuli, perception by birds, 79–80
- Schubert's Unfinished Symphony, 279
- Second effect of pitch shift, 185
- Secondary auditory cortex, pitch center, 155ff  
     regularity representation, 155
- Seebeck, missing fundamental pitch, 181
- Selective listening, mechanisms, 116–117
- Self-organizing maps (SOMs), neural net, 338ff
- Sensory-driven models of perception, 306–307, 308ff

- Sequential grouping of speech by pitch, 299–300
- Sequential streaming, 296ff
- Sharpening of tonotopic pattern, pitch theory, 186–187
- Shepard tones, 322, 328
- Siebert, pitch theory, 189
- Sinusoidally amplitude-modulated (SAM) sounds, and pitch perception, 63, 104
- Sound level, effect on frequency difference limen, 10–11
- Sound source segregation, and cochlear implants, 268
- starling, 88
- Sparse imaging, BOLD response, 149
- Spatial mapping, and pitch theory, 217–218
- Specific energies, pitch theory, 183
- Spectral models of pitch, 152–153
- brain imaging, 155
- Spectral pattern, pitch perception, 170
- Spectral resolvability, 22
- Spectral/spatial mechanisms, frequency discrimination, 235ff
- Speech, changes in F0, 295
- Speech perception, and altered pitch perception, 266
- and competing sounds, 267
- and hearing loss, 254ff
- Speech perception deficits, and pitch perception, 254
- Spontaneous discharge rate, see Spontaneous rate
- Spontaneous rate, auditory nerve, 100–101
- cortical neurons, 130
- effect of background noise, 110
- frequency discrimination, 104
- SR, see Spontaneous rate
- Starling, see also *Sturnus vulgaris*
- absolute vs. relative pitch perception, 83
- auditory streaming, 87–88
- missing fundamental pitch, 66
- musical chord discrimination, 81
- octave generalization, 82
- pitch perception, 83
- pure tone frequency generalization, 61
- timbre perception, 66
- Starting phase discrimination, perception by nonhuman animals, 79–80
- Starting phase effects, nonhuman animals, 78–79
- Stellate cells, inferior colliculus, 127
- Stimulus generalization, and discrimination, 58ff
- Stimulus generalization paradigms, pitch perception in animals, 59–60
- Stimulus regularity, firing rate, 152
- Streaming, difference in localization, 298
- nonhuman animals, 87ff
- Stretched string, as model for pitch, 182
- Strobed temporal integration (STI), pitch theory, 197
- Sturnus vulgaris* (European starling), see also Starling
- pure tone frequency generalization, 61
- Summary autocorrelation function (summary ACF), 194–195
- Suppression, pitch theory, 187
- Sylvian fissure, superior temporal plane, 153
- Synchronization index, 103–104
- brain imaging, 152
- Synthetic listening, 16ff
- virtual pitch, 184
- Taeniopygia guttata* (zebra finch), see also Zebra finch
- detection of harmonic sounds, 70–71
- Target chord processing, fMRI, 333
- Tartini tones, beat frequency, 201
- Taylor series, combination tones, 202
- Templates, missing fundamental pitch, 183
- Temporal and spectral representations, orthogonality, 162
- Temporal codes, and cochlear implants, 261
- pitch, 11ff
- Temporal context, and pitch, 317–319
- Temporal information, primary-like units, 114
- Temporal integration and resolution, measures of, 40ff
- modulation sensitivity, 40–42
- pitch, 40ff
- pitch theory, 204ff

- Temporal mechanisms, frequency discrimination, 235ff
- Temporal models, and brain imaging, 155
  - pitch, 108
  - residue pitch, 105
- Temporal pattern, pitch perception, 171
- Temporal processing, chopper units in cochlear nucleus, 118–119
- Temporal regularity, and MEG, 162–163
  - pitch strength, 156
  - and rate coding, 150ff
- Temporal resolution, pitch of single tones, 102–104
- Temporal theory, early developments, 189ff
  - pitch models, 189ff
- Temporal-place representation, cochlea, 107
- Terhardt, pitch model, 184
- Thalamus, projections to auditory cortex, 133
- Timbre, perception of, 291ff
  - and pitch, 212ff
- Time domain processing, rippled noise pitch, 151
- Timing cues, cochlear implants, 260ff
- Tone height, and pitch, 176–177
- Tonotopic organization, inferior colliculus, 127–128
- Tonotopic organization of cortex, fMRI study, 154–155
  - PET study, 153–154
- Tonotopic projection, early theory, 179
- Tonotopy, auditory cortex, 130–131, 153ff
  - brainstem, 153
  - cochlear nucleus, 112
- Top-down processes, 306, 307, 308ff, 334
  - neurophysiological studies, 330
- Training, effects on F0DL, 39
- Traveling wave, propagation time altered, and cochlear damage, 237
  - pitch perception, 235ff
- Tursiops truncatus* (Atlantic bottlenosed dolphin), see also Dolphin
  - rippled noise perception, 75
- Unconscious inference, Helmholtz, 184
- Unresolved harmonics, 20ff, 29ff
  - integration mechanism, 46–47
  - phase effects, 32–33
  - pitch mechanisms, 4, 36–38, 236, 287
- Unresolved vs. resolved harmonics, 20ff
- VCN, response to iterated rippled noise, 126
- Ventral cochlear nucleus, see Cochlear nucleus, VCN
- Vibrato, 40
- Virtual pitch, Terhardt, 184
- Visual perception, 308ff
- Vowel pairs, identification, 294
- Vowels, discrimination with F0 differences, 291–202
  - representation in auditory nerve, 110
- Waveform, envelope, 32ff
  - fine structure, 32ff
- Western harmony, 322–324
- Western music, cognition, 306ff
  - pitch classes, 313–314
- Western pitch structures, modeling, 338ff
- Western tonal music, perception, 306ff, 311ff
- Wever and Bray, volley theory, 191
- Wiener–Khinchine theorem, 201
- Wightman, pitch model, 183–184
- Yodeling, 299
- Zebra finch, see also *Taeniopygia guttata*
  - missing fundamental pitch, 71
  - phase effects of harmonic complexes, 78
  - profile analysis, 71
  - starting phase discrimination, 79–80