

Mari Riess Jones
Richard R. Fay
Arthur N. Popper
Editors

(1) **Music Perception**



Springer Handbook of Auditory Research

For other titles published in this series, go to
www.springer.com/series/2506

Mari Riess Jones • Richard R. Fay
Arthur N. Popper
Editors

Music Perception

 Springer

Editors

Mari Riess Jones
The Ohio State University
Columbus, OH 43210
USA
and
University of California Santa Barbara
Santa Barbara, CA 93111
USA
jones.80@osu.edu

Richard R. Fay
Loyola University of Chicago
Chicago, IL
USA
rfay@luc.edu

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

ISBN 978-1-4419-6113-6 e-ISBN 978-1-4419-6114-3
DOI 10.1007/978-1-4419-6114-3
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2010931281

© Springer Science+Business Media, LLC 2010

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, LLC, 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 on acid-free paper

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

Series Preface

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

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

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

Richard R. Fay, Chicago, IL
Arthur N. Popper, College Park, MD

Volume Preface

This volume presents an overview of a relatively new field of psychoacoustic and hearing research. The field involves perception of musical sound patterns, and this is considered in a set of chapters that reflect the current status of scientific scholarship related to music perception. Each chapter aims at synthesizing a range of findings associated with each of several major research areas in the field of music perception. These chapters have been crafted to present conceptual, but not necessarily exhaustive, reviews of research addressed to the major issues in the field of music perception.

In Chapter 2, Patterson, Gaudrain, and Walters introduce the reader to theory and research concerned with the basics of musical note perception; they offer a new approach to the meaning and measurement of pitch and timbre. Krumhansl and Cuddy (Chapter 3) focus on relative pitch in musical events and the way it conveys tonal relationships to listeners. This is followed by Chapter 4, authored by Trainor and Corrigan, who discuss acquiring musical sensibilities, which are often customized within a given musical culture and present new findings using listeners drawn from many different age ranges and cultures. In Chapter 5, Schellenberg and Hunter lay out the theoretical and methodological complexities presented by studying various responses to music with a special focus on the way musical passages can elicit emotions. In Chapter 6, McAuley considers basic distinctions among tempo, meter, and rhythm as well as relevant methodological concerns. This is followed by Chapter 7 by Large, who adapts the framework of dynamical systems to describe neurodynamical models of pitch perception and musical listening and to explain musical universals. Finally, in Chapter 8, Halpern and Bartlett provide an overview of people's ability to recognize melodies and consider the ability of people of different ages to recognize familiar and novel melodies.

As in other SHAR volumes, chapters in this volume are complemented by earlier chapters and volumes in the series, and particularly in those volumes that cover perception of sound, in general, and speech, in particular. These include chapters in Volume 3 (*Human Psychophysics*), Volume 24 (*Pitch*), and Volume 29 (*Auditory Perception of Sound Sources*). In addition, many of the ideas regarding music perception are paralleled in Volume 18 (*Speech Processing in the Auditory System*).

Mari Riess Jones, Santa Barbara, CA
Richard R. Fay, Chicago, IL
Arthur N. Popper, College Park, MD

Contents

1 Music Perception: Current Research and Future Directions	1
Mari Riess Jones	
2 The Perception of Family and Register in Musical Tones	13
Roy D. Patterson, Etienne Gaudrain, and Thomas C. Walters	
3 A Theory of Tonal Hierarchies in Music.....	51
Carol L. Krumhansl and Lola L. Cuddy	
4 Music Acquisition and Effects of Musical Experience	89
Laurel J. Trainor and Kathleen A. Corrigan	
5 Music and Emotion	129
Patrick G. Hunter and E. Glenn Schellenberg	
6 Tempo and Rhythm.....	165
J. Devin McAuley	
7 Neurodynamics of Music.....	201
Edward W. Large	
8 Memory for Melodies.....	233
Andrea R. Halpern and James C. Bartlett	
Index.....	259

Contributors

James C. Bartlett

School of Behavioral and Brain Sciences, University of Texas at Dallas,
Richardson TX 75083, USA
jbartlet@utdallas.edu

Kathleen A. Corrigan

Department of Psychology, Neuroscience & Behaviour, McMaster University,
Hamilton ON, Canada, L8S 4K1
corrigan@mcmaster.ca

Lola L. Cuddy

Department of Psychology, Queen's University, Kingston, Ontario,
K7L 3N6, Canada
Lola.Cuddy@queensu.ca

Etienne Gaudrain

MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, CB2 7EF,
Cambridge, U.K.
etienne.gaudrain@mrc-cbu.cam.ac.uk

Andrea R. Halpern

Psychology Department, Bucknell University, Lewisburg, PA 17837, USA
ahalpern@bucknell.edu

Patrick G. Hunter

Department of Psychology, University of Toronto at Mississauga,
Mississauga ON, L5L 1C6, Canada
patrick@psych.utoronto.ca

Mari Riess Jones

Departments of Psychology, The Ohio State University and University of
California, Santa Barbara, Santa Barbara, CA 93111, USA
jones.80@osu.edu

Carol L. Krumhansl

Department of Psychology, Cornell University, Ithaca, NY 14853, USA
clk4@cornell.edu

Edward W. Large

Center for Complex Systems & Brain Sciences, Florida Atlantic University,
Boca Raton, FL 33431, USA
large@walt.ccs.fau.edu

J. Devin McAuley

Department of Psychology, Michigan State University, East Lansing,
MI 48823, USA
dmcauley@msu.edu

Roy D. Patterson

Centre for the Neural Basis of Hearing, Department of Physiology,
Development and Neuroscience, University of Cambridge,
Cambridge CB2 3EG, UK
rdp1@cam.ac.uk

E. Glenn Schellenberg

Department of Psychology, University of Toronto at Mississauga,
Mississauga, ON, Canada L5L 1C6
g.schellenberg@utoronto.ca

Laurel J. Trainor

Department of Psychology, Neuroscience & Behaviour, McMaster University,
Hamilton, ON, Canada L8S 4K1
LJT@mcmaster.ca

Thomas C. Walters

Google Inc., 1600 Amphitheatre Parkway, Mountain View, CA, 94115, USA
tomwalters@google.com

Chapter 1

Music Perception: Current Research and Future Directions

Mari Riess Jones

1.1 Introduction

In the rich tradition of the *Springer Handbook of Auditory Research* series, the present volume continues to provide an accessible overview of a relatively new field of psychoacoustic and hearing research. The field of central interest in this volume involves perception of musical sound patterns. In keeping with the goals of the Handbook, it presents a set of chapters that reflect the current status of scientific scholarship related to music perception. Each chapter aims at synthesizing a range of findings associated with each of several major research areas in the field of music perception. Thus, topics central to this field involve pitch perception, responses to harmony/tonality, tempo and rhythm, emotion and music, and finally melody recognition. These chapters have been crafted to present conceptual, but not necessarily exhaustive, reviews of research addressing the major issues in the field of music perception. The dominant issues, hypotheses, and theories that drive research within each topical area are presented along with relevant experimental findings. The aim is to introduce this growing research area to new investigators and established researchers unfamiliar with music perception scholarship.

1.2 Overview of this Volume

This volume opens appropriately with very basic questions that concern pitch perception when the sounds involved are pulse resonant ones generated by various musical instruments including the human voice. Patterson, Gaudrain, and Walters (Chap. 2) introduce the reader to important theory and research concerned with

M.R. Jones (✉)

Departments of Psychology, The Ohio State University and University of California,
Santa Barbara, Santa Barbara, CA 9311
e-mail: jones.80@osu.edu

basic aspects of musical perception. They also outline an original approach to this topic that attempts to rethink the meaning and measurement of pitch and timbre. They raise fundamental questions about why we perceive various instruments as belonging to musical families of sound sources. In this they explore aspects of musical instruments that function as source and filter and how resulting sound properties are respectively scaled so the listeners can reliably identify families of instruments. The authors address complicated and historically poorly understood issues that surround a formalization of timbre. Chap. 2 creatively integrates a number of novel findings with bold ideas about pitch perception, suggested by the influential Auditory Image Model (AIM).

Krumhansl and Cuddy (Chap. 3) also address pitch, but at a different level than described in Chap. 2. The major focus of their chapter is on relative pitch in musical events and the way it conveys tonal relationships to listeners. Krumhansl and Cuddy present a complete portrait of a major theory of tonality perception, one that drives a majority of the research on this topic. The theory is Krumhansl's tonal hierarchy theory (Krumhansl 1990), and it is lucidly described in this chapter. The authors succeed in conveying the richness and diversity of empirical research that has developed on the topic in recent years. Among other things, this research considers how listeners respond to musical keys and to various sound sequences, which include ones that reinforce a primed key (e.g., C Major) and those that imply a modulation to another key (e.g., G Major) as well as musical sequences that contain an outright violation of a primed key. These and related topics are considered as they relate to the tonal hierarchy construct. In addition, the overview presented by Krumhansl and Cuddy outlines the basic features of this approach by connecting these features to general psychological principles. The latter involve cognitive prototypes, namely salient constructs and referent points around which instances of a category are organized. The authors show that tonality involves listeners' tendencies to rely on certain tones, such as the keynote, as cognitive reference points, about which other scale tones are organized. In addition, they provide links to statistical learning principles to suggest how people acquire the ability to infer tonal relationships from an arbitrary melody. A wealth of empirical findings is presented to document the dominant role that an acquired sense of tonality plays in a listener's response to music.

The issue of how people acquire the skills that render them keen musical listeners is a pervasive and fascinating one. It recurs in other chapters as well. Yet the ability to perceive a particular musical key or a special cadence or to determine a particular metrical frame is as mysterious to those who do not have these skills as it seems effortless to those who have them. But for scientists who enjoy studying the mysterious, this is a challenge that begins with an infant's early responses to music and extends to understanding the impact of a cultural environment throughout the lifespan. With respect to acquiring musical sensibilities, which are often customized within a given musical culture, Trainor and Corrigan (Chap. 4) provide an engaging essay. The authors are among the leaders in field of developmental research on music perception. In their chapter, they present new and interesting findings gathered from listeners drawn from many different age ranges and cultures. In particular,

Trainor and Corrigan present interesting discussions related to the role of experience in developing musical skills. For instance, although very early in life infants show sensitivity to differences in tone consonance, they appear to acquire a true sense of key only several years later. The authors tackle as well differences in perception based on experience in children and adults including comparisons of listening skills of musicians and nonmusicians. In short, this is a rich chapter that nicely integrates information from a range of methodologies (functional magnetic resonance imaging [fMRI], event-related potential [ERP], etc.) as well as data from listeners drawn from a range of different age groups and musical skill levels.

It is difficult to think of responses to music without including a prominent role for affect and the way musical passages can elicit emotions. In a thorough overview of the research on music and emotion, Schellenberg and Hunter (Chap. 5) detail the theoretical and methodological complexities in studying this topic. How does one go about measuring emotions? This question is a theme that runs throughout the chapter as different theoretical approaches are accompanied by different answers. Moreover, other questions quickly follow: Can emotional responses to music be explained using a two-dimensional model of emotion (i.e., dimensions of valence and arousal)? Do people simply perceive that a piece of music is designed to be sad or happy or do they actually partake of one or another of these emotions when listening? Does listening to music (e.g., Mozart) really make one smarter, as some contend, or has such research been misinterpreted? The authors present an overview of the tools and methods that psychologists have brought to bear on these questions and they thoroughly discuss pros and cons of issues arising from such questions.

In many respects, the quality of a musical performance is judged by its timing. Is the tempo correct? Is the rhythm appropriately conveyed? From the perspective of a composer, tempo and rhythm/meter frame a musical event and are an integral part of the message. From the perspective of a listener, these timing properties not only shape the musical message, but they can also afford to listeners certain rates at which they may pace their attending to track a musical piece in real time. But of course to answer the foregoing questions regarding the appropriateness of a particular tempo or rhythm, listeners must be able to render judgments about the relative timing of a performance, after the fact (e.g., "Is it too slow or too fast?"). Bringing these tasks into the laboratory has entailed the inevitable simplification of musical events, often to monotone and/or isochronous sequences. McAuley (Chap. 6) presents a clear and engaging picture of how this is achieved and the issues confronting current research on musical time. Basic distinctions among tempo, meter, and rhythm are provided, as are relevant methodological concerns. For example, some tasks require that people merely judge some temporal property (tempo, meter) whereas others require that they reproduce the timing of entire sequences (e.g., in motor tapping tasks). McAuley also offers a lucid discussion of major theoretical topics that characterize this area of research. Key distinctions between interval timing theories and entrainment (beat-based) timing theories are outlined. Thus, interval timing theories maintain that judgments of tempo can be best described using a statistical analysis of the different time intervals in a musical piece such that the mean and standard deviation of a pattern's time intervals determine judgments about whether one tempo

is fast or slow (relative to another pattern). Alternatively, entrainment approaches suggest that listeners synchronize their attention to recurrent periods within an unfolding event such that stimulus tempo is conceived as driving rhythm where attending oscillations are depicted as driven rhythms. McAuley's own research on tempo perception is leading the way to clarifying these issues.

In his chapter on neurodynamics of music (Chap. 7), Large discusses some of the formalisms and models that underlie the dynamics of entrainment. In addition, at a more general level, Large observes that although music shares with language universality in that both are present in some form in all cultures, music is unlike language in that it is mainly self-referential. That is, whereas language routinely contains referents that encourage a focus on aspects of an external world, music exhibits an independence from external reference. Large shows how both the universality and the independence of music can be accommodated using general principles of neural functioning and models based on dynamical systems concepts. He addresses issues concerned with time scales and resonance, but here we find an approach to resonance that is fundamentally nonlinear. In addition, other constructs basic to nonlinear dynamics such as oscillations (of neural populations), rhythmic bursting, and neural synchrony are introduced in high-level dynamical models, that is, canonical models. Large illustrates ways in which these constructs can be elegantly adapted to explain well-established phenomena such as pitch shift data, certain aspects of tonality perception, and meter perception.

Finally, Halpern and Bartlett (Chap. 8) offer a wealth of findings on the ability of people of different ages to recognize familiar and novel melodies. Although melodies typically reflect invariance in relative pitch (i.e., they can be transposed to different octaves and keys, yet retain their identity), emerging research indicates an intriguing role for absolute pitch in melody recognition. Further, even presenting a melody at its original tempo appears to aid recognition memory. Also prominent among the various determinants of good melody recognition is the degree to which a melody is nameable. One paradoxical issue they highlight concerns the fact that although learning a novel melody is difficult, in general memory for melodies is quite good for most people even into their old age. Perhaps because issues associated with structural determinants of melody perception are formidable, most recent research surrounding the topic of melody has focused on memory for melodies and determinants of recognition memory. As Halpern and Bartlett document, people are best at recognizing melodies that are perceived as familiar ones and are nameable.

1.3 Future Directions

As suggested, the field of music perception research is a relatively young area of inquiry when considered in light of other major areas of research involving psychoacoustics, visual perception, memory, or even speech perception. To be sure, classic works by Seashore (1938), Francés (1958), and Fraisse (1963) came around

the middle of the last century. But these approaches to music were unusual efforts even for their time, hence remarkable for their pioneering qualities; it took time for them to “take root.” And, although important groundwork in pitch perception was laid down by laboratories in the Netherlands, renowned for their psychoacoustic discoveries using single complex tones, contemporary researchers in other countries in the latter part of the last century also prefigured the growth of current research seen today. In addition, pioneers such as Diana Deutsch, Roger Shepard, Al Bregman, and Dirk Povel, among others, began to consider how people perceive and remember whole sequences of tones. Significant breakthroughs in those days came with demonstrations that people responded to harmonic and melodic relationships between tones within longer sequences and they could also respond systematically to metrically simple versus complex sequences. Against this background, the present chapters make clear that music perception research has progressed much beyond these early breakthroughs to include studies that examine precisely how people may perceive and acquire skills in listening to musical sound patterns. So where will this take the field in the future?

It is possible only to speculate on paths that may be taken. However, a case can be made that future progress can be anticipated along three paths. One path is methodological; this broadly covers both newer investigative tools (fMRI, EEG, etc.) and use of more familiar statistical tools. Another path involves development of new hypotheses, models, and techniques that will address musical complexities. A third path may lead to revival of interest in understanding melody perception.

1.3.1 Methodological Progress

Certainly the direction in which much research in psychology is moving invites increasingly greater reliance on neuroscience methodologies. Research in music perception is no exception. This is an exciting time. In many of the chapters of this book, recent findings from fMRI, EEG, and positron emission tomography (PET) have been incorporated into discussions of a particular phenomenon and these often advance understanding. This path portends an emerging capacity in the field, as a whole, for identifying regions and circuits in the brain responsible for attending to and perceiving specific aspects of music. At the current stage of this daunting undertaking, questions of interest often revolve around identifying brain regions specialized to respond to major components of music conventionally identified as tonal structure, melody, or rhythm. Although it is possible that the brain will reveal neatly segregated pathways and regions specific to these traditional musical categories, it is equally likely that the picture will turn out to be much more complicated and challenging for future researchers.

The complications arise not only from the obvious fact that mapping brain activity itself broaches unknown territory, but also from the fact that this unknown territory itself will likely be highly complex. Adding to the intricacies, complications also attend the stimuli that are commonly used to study brain activity. Typically

these are musical patterns that are often quite complex in their own right. One compelling reason for using composed music, with a degree of complexity, as stimuli is that these sound patterns reflect the true complexity of “real music.” In other words, stimuli must be ecologically valid. This rationale cannot be discounted. At the same time, it should not be embraced without constraint. That is, it is important to acknowledge that valid musical events notoriously covary many of the major components of music (melody, harmony, tempo, rhythm). As a result, a clear cautionary note can be sounded regarding interpretations of findings from brain imaging/scan research that relies solely on melodies sampled from the corpus of composed melodies. To the degree these stimuli contain covariations of melodic, harmonic, and rhythmic structural components, they can present as serious cases of confounded variables. In the future, it is likely that calls for better controls over both experimental designs and stimulus structure will be heeded in applying new technologies.

Another methodological issue is a statistical one. It is not unrelated to the preceding issues in that when ecologically valid composed music serves as stimuli for a psychological investigation, resulting statistical analyses are often correlational. For instance, listeners’ judgments about a musical property, which may involve ratings of goodness, closure, clarity and so forth, might be subjected to a correlational analysis to ascertain if judgments are reliably associated with differences in musical style or tempo or mode, etc. This methodological approach is understandable, particularly when the stimulus pattern can vary along many different dimensions. In a new area of research, the first step in a discovery process is typically to cast a broad net to ascertain what features of a complex situation might be critical to investigate: Does X change with Y? In this respect, correlational data can and should lead the way and help to formulate initial hypotheses.

Nevertheless, common cautions regarding correlational research are worth repeating. As the field of music research matures, two points should be made about a continuation of heavy reliance on correlational designs. Experimentally speaking, neither of them is novel. The first point is that correlational evidence, finally, does not support inferences about causal relationships between variables X and Y. The second point is that when correlation is used in conjunction with specifically selected musical excerpts, the preselection of materials can distort resulting correlation coefficients, leading to over- or underestimates of the true relationship between variables. Therefore, as this broad research area advances, it is to be expected that correlational designs will be supplemented by, or give way to, more rigorous designs that incorporate true experimental manipulations. In the future, this will allow more reliable causal inferences as well as the development of more specific models of music perception.

1.3.2 Musical Complexity

The preceding section has raised certain methodological issues surrounding selection of “real music” as stimuli for investigations in music perception. One of

these issues concerns the fact that in real music many different components naturally covary (i.e., harmony, melody, rhythm, tempo, etc.). Real music is complex. One musical component can change over time within a single piece in ways that highlight certain components but not others. Rhythm can, for example, enhance a melodic line in some phrases and obscure it in others. It is this complexity that makes music interesting and that motivates researchers to select real music as stimuli for experimental investigation rather than dull contrived laboratory music. At the same time, it is evident that singular reliance on stimuli sampled from the corpus of composed music poses limitations on an experimenter's understanding of precisely what is actually determining a listener's response.

The ultimate goal is to embrace a strategy that allows some insight into musical complexities while retaining control over variables that contribute to this complexity. To supplement correlational analyses, this may entail new adaptations of classic experimental methodologies that encourage holding constant one variable while systematically manipulating another. One strategy is to consider creative inclusions of control stimuli that experimentally eliminate or modify certain features of interest in musical patterns sampled from a known corpus. Another strategy is to abandon reliance on complex, but ecologically valid, musical events and embrace traditional experimental methods that do simply vary one property of sound pattern, holding others constant. Of course, there already is a significant body of current research that does just this. This methodology comes at a cost, however. Musical complexity and ecological validity are both greatly reduced. That is, when an initially complex musical stimulus is stripped of variations in all components except those associated with a single component of interest, it becomes simpler and more manageable but it loses some musical validity. For instance, when studying melody perception, only the melodic (pitch) component is varied while the rhythmic component is held constant, that is, no timing variations are introduced. Similarly, when examining rhythm perception, typically the melodic component of a rhythmic pattern is removed, such that sequences comprise tones of a single pitch. In fact, as noted earlier, many of the traditional topics of interest in music perception reflect this methodological approach. It is even evident in chapter titles in this volume: tonality/pitch, rhythm/meter, and melody. Typically, studies on the topic of tonality or harmonic progressions vary successive pitch relationships, while holding constant rhythm, meter, and tempo, for example. Conversely, when examining temporal aspects of rhythm and/or meter, it is common to hold constant pitch relationships by employing monotone sequences in which the same pitch is repeated throughout. Consequently, this strategy results in rigorous experiments that allow for reliable causal inferences about the effects of an important component of music. These studies have served an important goal in amassing the current knowledge. But this gain in knowledge comes at the price, which is a loss of musical complexity and limits on the generality of these inferences. Often these laboratory stimuli are fairly simple and not aesthetically interesting. Moreover, they do not provide insight into how different musical component may interact, if they do.

A challenge for the future will be to move beyond simple experimental designs to ones that allow creative investigations of how different musical components

work together or play off against one another. For instance, there is some evidence that rhythm affects both melody perception (e.g., Kidd et al. 1984; Jones and Ralston 1991; Jones et al. 2006) and perceived tonal endings (e.g., Boltz 1989; Schmuckler and Boltz 1994). How can such findings be explained? These questions have been touched on in the research discussed in chapters of the volume authored by Krumshank and Cuddy (Chap. 3), McAuley (Chap. 6), and Halpern and Bartlett (Chap. 8), but complete answers will come only with more research and new models. This endeavor remains challenging in part because it involves striking the right balance between experimental control and real musical complexity. One solution is to begin to incorporate two factor designs that allow for systematic manipulations of each of two types of musical components, thereby creating more complicated music-like sequences. A few examples of this within the literature involve combinations of melody and rhythm; these have led to interesting debates about the respective roles of different kinds of accents (e.g., melodic, rhythmic) and about the degree to which melody and rhythm are psychologically dependent in a given task (for a brief review see Ellis and Jones 2009).

1.3.3 *Melody Perception*

In his 1959 book, *Productive Thinking*, Max Wertheimer (1959) speculated on the relational properties of melodies. Since those days, melody perception has fascinated psychologists. Early Gestalt theorists saw various relationships between successive sounds in common melodies as realizations of innate laws of good organization (pragnanz) that governed the perception of grouping structures in the world. The fact that the most frequently occurring melodic intervals, throughout world cultures, are small (i.e., less than three semitones in pitch distance) was seen as validation of the law of proximity, whereas the prevalence of ascending and descending pitch movements appears to confirm the Gestalt law of continuity (Dowling and Harwood 1986, pp. 155–160). Echoes of the Gestalt approach to melody perception are quite clear even today in the voluminous work of two major figures in the field, namely Narmour (1990) and Bregman (1990). Whereas Narmour points to supporting evidence in analyses of real music, Bregman finds validation for Gestalt principles in experimental assessments of listeners' responses to simplified music-like sequences that are consistent with Gestalt principles. Although Bregman argues that innate Gestalt principles affect perception, he contends that they do not do so by determining melodic expectancies which, he argues, are essentially learned. On the other hand, Narmour, who also appeals to innate Gestalt rules, maintains these principles do determine listeners' melodic expectancies, as well as perception. He not only relies on the Gestalt principles of proximity and continuity, but also argues for expectancies based on similarity, and pitch trajectory reversals (i.e., pitch trajectory reversals) as well. However, Schellenberg (1997) elegantly demonstrated that expectancies about melodic endings, at least for fairly simple melodies, can be parsimoniously described using a relatively small number of Gestalt rules.

Finally, although Gestalt principles are useful descriptors of salient pattern relationships, they do not offer a complete explanation of melody perception. Indeed, not all of those who focused on relational properties of melodies can be considered Gestalt theorists. Roger Shepard concentrated on formalizing efficient geometric paths in mental spaces that listeners use in responding to salient pitch relationships. For instance, abstract pitch relationships, critical to Western musical scales, might assume the mental form of a double helix pattern (Shepard 1982). W. Jay Dowling showed how sensitive listeners are to melodic contour and to systematic atonal transformations of melodic relationships among pitch intervals, such as transposition, retrograde transformations (Dowling and Fujitani 1971). Diana Deutsch argued for economical pitch encoding rules (Deutsch and Feroe 1981), and Mari Jones used mathematical group theory to describe perceptually compelling symmetry relationships that often characterize melodic figures, relationships that can be either perceptually preserved or broken (i.e., into auditory pattern streams) depending on the tempo or rhythm of a melody (Jones 1976, 1981; Boltz et al. 1985; Boltz and Jones 1986).

Recent years have witnessed a decline in theoretical and empirical work on the role of melodic relationships in perception and perceptual learning, whether or not these relations are appropriately described by Gestalt principles. This explains why there is no chapter in this volume that is simply entitled *melody perception*. There are many reasons for the waning of research on the perception of melodic relationships. One is simply a growing fascination with understanding acquired listening skills that seem to underlie listeners' evident sensitivity to musical tonalities within different cultures and their accompanying sensitivity to specific harmonic progressions that may be implied by even a single melodic line. That is, it is clear that Gestalt rules do not really address tonal relationships in music, yet as both Krumhansl and Cuddy (Chap. 3) and Large (Chap. 7) illustrate, in different ways, tonality is a major factor in perception of melodic relationships.

Another reason for a decline in interest in melody perception, as such, is found in the (possibly mistaken) tendency to assume that Gestalt rules have successfully explained innate aspects of melody perception. This is debatable, I think, as suggested earlier. Nevertheless, this belief contributes to the interest in learning as a factor in explaining why certain compelling relationships appear to dominate percepts in musical patterns that Gestalt principles fail to address; the latter include consonance versus dissonance, tonality, and key relationships, among others. This growing interest in learning has been spurred by the gathering influence of statistical learning theory which revives age-old ideas of associationism, now based on temporal distributions of conditional probabilities. These probabilities reflect dependencies among two and three successive items within simplified sequences of tones or syllables (Saffran et al. 1999; Saffran 2003). Saffran and colleagues have shown that infants and adults can learn segmentations of both music-like and language-like sound sequences based on such contingencies, seemingly regardless of relational features among these tones. The growing impact of statistical learning theory is evident in many chapters of this volume. It explains why Halpern and Bartlett (Chap. 8) focused on the acquisition and memory for melodies instead of melody perception.

Although it is undeniable that learning is responsible for listeners' sensitivities to various sound patterns, including musical ones, it is nonetheless possible to take issue with the claim that melody perception is entirely explained by a statistical description of the underlying learning process. Certainly it is telling that when presented with melodic sequences, infants are sensitive to and appear to favor relative pitch over absolute pitch in melodies without much exposure to these sequences, as Trainor and Corrigan observe in their chapter (Chap. 4). In addition, there is evidence of very early reliance on rhythmic properties; for example, newborns distinguish among different languages by apparently relying on rhythmic properties, evidencing a bias for music-like periodicities (Nazzi et al. 1998, 2000). Adults also rely heavily on rhythmic properties to differentiate melodies; they have difficulty identifying a learned melodic sequence if its original rhythm changes, even when temporal segmentations and statistical pitch properties are unchanged (e.g., Jones and Ralston 1991). In other words, it seems clear that from an early age, people rely on more than simply pitch dependencies in listening to melodic aspects of music and even speech.

Future research on these and other issues associated with melody perception may witness a revival due, in part, to increasing interest in the possible overlap between language and music. For many years speech has been considered a special domain, with speech perception putatively determined by special articulatory motor codes (Liberman et al. 1967). But this longstanding view has recently been challenged by an array of research showing that music-like (nonspeech) melodic contexts can affect simultaneous and subsequent perception of speech/language (e.g., Holt and Kluender 2000; Holt et al. 2000; Bigand et al. 2001; Koelsch et al. 2005; Dilley and McAuley 2008). Others have demonstrated domain overlaps by showing that harmonic musical contexts influence phoneme monitoring (Bigand et al. 2005). Further, both behavioral and ERP data reveal that young children with musical training are better at detecting weak pitch incongruities in both musical and linguistic sound patterns than their counterparts without musical training (Magne et al. 2006); moreover fMRI studies show that musical structure processing activates regions that are also responsible for language processing (e.g., inferior frontal regions and Broca's area; Tillman et al. 2006). Still others have noted a number of parallel structural features in speech prosody and melody/rhythmic figures (e.g., Jones and Boltz 1989; Schön et al. 2005; Quené and Port 2005; Patel 2008; Slevc et al. 2009), parallels that inevitably invite hypotheses about common mechanisms responsible for perception of music and speech. Indeed, in a sweeping overview of the biology and evolution of music, Fitch (2006) convincingly concludes that the structural "design features" of language and music (both broadly defined) across species suggest an overlap of core aspects of these two domains. It is clear that the field is changing in exciting ways.

Increasingly, recent research that is beginning to explore domain overlaps, namely commonalities between speech and music, involves some covariation of melodic with rhythmic features. As a result, issues related to stimulus complexity and to the importance of careful experimental control of component properties (e.g., melody and rhythm) will continue to be critical ones as this research area

grows in the future. It is also evident that perceptual learning, beginning early in life, will continue to be an essential part of the story of melody perception. However, many factors critical to perceptual learning of musical (and speech) patterns remain to be identified. At least with regard to melodic perception/recognition, future theorizing may lie somewhere in between a *tabula rasa* approach implied by a statistical learning view and full endorsement of nativistic principles implied by Gestalt grouping principles. Finally, as Halpern and Bartlett (Chap. 8) note in their conclusions, it is possible that a revival of interest in configurational aspects of melodic sequences will emerge, if only to discover constraints on the kinds of melodic relationships that speed or slow a listener's ability to learn a melodic sequence.

References

- Bigand E, Tillman B, Poulin, B, D'Adamo DA (2001) The effect of harmonic context on phoneme monitoring in vocal music. *Cognition* 81:B11–B20.
- Bigand E, Tillman B, Poulin-Charronat B, Manderlier D (2005) Repetition priming: is music special? *Q J Exp Psychol* 58:1347–1375.
- Boltz M (1989) Perceiving the end: Effects of tonal relationships on melodic completion. *J Exp Psychol Hum Percept Perform* 15:749–761.
- Boltz M, Jones MR (1986) Does rule recursion make melodies easier to reproduce? If not, what does? *Cogn Psychol* 18:389–431.
- Boltz M, Marshburn E, Jones MR, Johnson WW (1985) Serial-pattern structure and temporal-order recognition. *Percept Psychophys* 37:209–217.
- Bregman A (1990) *Auditory Scene Analysis*. Cambridge, MA: MIT Press.
- Deutsch D, Feroe J (1981) The internal representation of pitch sequences in the form of hierarchies. *Psychol Rev* 88:503–522.
- Dilley LC, McAuley JD (2008) Distal prosodic context affects word segmentation and lexical processing. *J Mem Lang* 59:294–311.
- Dowling WJ, Fujitani DS (1971) Contour, interval, and pitch recognition in memory for melodies. *J Acoust Soc Am* 49:524–531.
- Dowling WJ, Harwood DL (1986) *Music Cognition*. London: Academic Press.
- Ellis R, Jones MR (2009) The role of accent salience and joint accent structure in meter perception. *J Exp Psychol Hum Percept Perform* 35:264–280.
- Fitch WT (2006) The biology and evolution of music: A comparative perspective. *Cognition* 100:173–215.
- Francés R (1958 translated by Dowling WJ 1988) *The Perception of Music*. Hillsdale, NJ: Lawrence Erlbaum.
- Fraisse P (translated by Leith J 1963) *The Psychology of Time*. New York: Harper & Row.
- Holt LL, Kluender KR (2000) General auditory processes contribute to perceptual accommodation of coarticulation. *Phonetica* 57:179–180.
- Holt LL, Lotto AJ, Kluender KR (2000) Neighboring spectral content influences vowel identification. *J Acoust Soc Am* 108:710–722.
- Jones MR (1976) Time, our lost dimension: toward a new theory of perception, attention and memory. *Psychol Rev* 83:323–355.
- Jones MR (1981) A tutorial on some issues and methods in serial pattern research. *Percept Psychophys* 30:492–504.
- Jones MR, Boltz M (1989) Dynamic attending and responses to time. *Psychol Rev* 96:459–491.

- Jones MR, Ralston J (1991) Some influences of accent structure on melody recognition. *Mem Cognit* 19:8–20.
- Jones MR, Johnston MJ, Puente J (2006) Effects of auditory pattern structure on anticipatory and reactive attending. *Cogn Psychol* 53:59–96.
- Kidd G, Boltz M, Jones MR (1984) Some effects of rhythmic context on melody recognition. *Am J Psychol* 97:153–173.
- Koelsch S, Gunter TC, Wittforth M, Sammler D (2005) Interaction between syntax processing in language and in music: an ERP study. *J Cogn Neurosci* 17:1565–1579.
- Krumhansl CL (1990) *Cognitive Foundations of Musical Pitch*. New York: Oxford.
- Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M (1967) Perception of the speech code. *Psychol Rev* 74:431–461.
- Magne C, Schön D, Besson M (2006) Musician children detect pitch violations in both music and language better than nonmusician children: behavioral and electrophysiological approaches. *J Cogn Neurosci* 18:199–211.
- Narmour E (1990) *The Analysis and Cognition of Basic Melodic Structures*. Chicago, IL: University of Chicago Press.
- Nazzi T, Bertoncini J, Mehler J (1998) Language discrimination by newborns: toward an understanding of the role of rhythm. *J Exp Psychol Hum Percept Perform* 24:756–766.
- Nazzi T, Jusczyk PW, Johnson EK (2000) Language discrimination by English-learning 5-month-olds: effects of rhythm and familiarity. *J Mem Lang* 43:1–19.
- Patel A (2008) *Music, Language and the Brain*. New York: Oxford University Press.
- Quené H, Port RF (2005) Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica* 62:1–13.
- Saffran JR (2003) Statistical language learning: mechanisms and constraints. *Curr Dir Psychol Sci* 12: 110–114.
- Saffran JR, Johnson EK, Aslin RN, Newport EL (1999) Statistical learning of tone sequences by human infants and adults. *Cognition* 70:27–52.
- Schellenberg EG (1997) Simplifying the implication-realization model of musical expectancy. *Music Percept* 14: 295–318.
- Schmuckler MA, Boltz MG (1994) Harmonic and rhythmic influences on musical expectancy. *Percept Psychophys* 56:31–325.
- Schön D, Gordon RL, Besson M (2005) Musical and linguistic processing in song perception. *Ann NY Acad Sci* 1060:71–81.
- Seashore CE (1938) *The Psychology of Music*. New York: McGraw Hill.
- Shepard RN (1982) Geometrical approximations to the structure of pitch. *Psychol Rev* 89:305–333.
- Slevc LR, Rosenberg JC, Patel AD (2009) Making psycholinguistic musical: self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychon Bull Rev* 16:374–381.
- Tillman B, Koelsch S, Escoffier N, Bigand E, Lalitte P, Friederici AD, vonCramon DY (2006) Cognitive priming in sung and instrument music: activation of inferior frontal cortex. *NeuroImage* 31:1771–1782.
- Wertheimer M (1959) *Productive Thinking*. New York: Harper & Row.

Chapter 2

The Perception of Family and Register in Musical Tones

Roy D. Patterson, Etienne Gaudrain, and Thomas C. Walters

2.1 Introduction

This chapter is about the sounds made by musical instruments and how we perceive them. It explains the basics of musical note perception, such as why a particular instrument plays a specific range of notes; why instruments come in families; and why we hear distinctive differences between members of a given instrument family, even when they are playing the same note. The answers to these questions might, at first, seem obvious; one could say that brass instruments all make the same kind of sound because they are all made of brass, and the different members of the family sound different because they are different sizes. But answers at this level just prompt more questions, such as: What do we mean when we say the members of a family produce the same sound? What is it that is actually the same, and what is it that is different, when different instruments within a family play the same melody on the same notes? To answer these and similar questions, we examine the relationship between the *physical* variables of musical instruments, such as the length, mass, and tension of a string, and the variables of auditory *perception*, such as pitch, timbre, and loudness. The discussion reveals that there are three *acoustic* properties of musical sounds, as they occur in the air, between the instrument and the listener, that are particularly useful in summarizing the effects of the physical properties on the musical tones they produce, and in explaining how these musical tones produce the perceptions that we hear.

The remainder of the introduction sets out the aspects of tone perception to be explained, namely, the perception of pitch, instrument family, and instrument register within a family. The second section describes the acoustic properties of tones as they pertain to music perception, and sets out some of the terminology used in the chapter. The third section explains the relationship between the physical

R.D. Patterson (✉)

Centre for the Neural Basis of Hearing, Department of Physiology,
Development and Neuroscience, University of Cambridge, Cambridge CB2 3EG, UK
e-mail: rdp1@cam.ac.uk

variables of tone production (length, mass, tension, etc.) and the acoustic variables observed in the sounds. The fourth section describes the internal representation of musical sounds in the auditory system to show how the acoustic properties of sound appear in the auditory representation of musical tones. The fifth and final section reviews the relationship between the acoustic variables of sound and the auditory variables of tone perception, and suggests how the standard definitions of pitch and timbre might be revised for use in discussions of the perception of musical tones and musical instruments.

2.1.1 *Pitch, Instrument Family, and Instrument Register Within a Family*

The chapter focuses on the sounds produced by the sustained-tone instruments of the orchestra and chorus, that is, the families of instruments referred to collectively as brass, strings, woodwinds, and voice. Table 2.1 shows four of the instruments in each of the families, ordered in terms of their size or their *register*. With just a small amount of training, most people can learn to identify these 16 instruments from a simple monophonic melody (van Dinther and Patterson 2006). With regard to family and register, the purpose of the chapter is to explain how auditory perception enables us to distinguish the main families and the different instruments within a family.

Imagine the sequence of tones you would hear if a trombonist, a cellist, a bassoonist, and a baritone vocalist would in turn produce the same tone, say C3 (the C below middle C on the keyboard). What is the “same” about the four tones is their pitch. What is different, and what allows us to distinguish the tones, is the distinctive timbres of the different instrument families. This is the traditional distinction between the *perceptual* variables, pitch and timbre. The pitch of a musical tone is effectively determined by the repetition rate of the sound. The sound waves produced by the sustained-tone instruments of the orchestra (brass, string, woodwind, and voice) are complex and their spectra are complex; nevertheless the tones are essentially periodic and the pitch that they produce is very closely related to the number of times that the sound wave repeats in the course of 1 s. This aspect of music perception is entirely straightforward for sustained-tone instruments. Psychoacousticians have developed models to explain how the auditory system extracts pitch from sound waves, and the models have become increasingly elaborate as they attempt to explain the pitches produced by exotic, computer-generated

Table 2.1 Sixteen common instruments illustrating four registers within each of four instrument families

Register/family	Brass	Strings	Woodwind	Voice
High	Trumpet	Violin	Soprano sax	Alto voice
Mid-high	Trombone	Viola	Alto sax	Tenor voice
Low-mid	French Horn	Cello	Tenor sax	Baritone voice
Low	Tuba	Contra bass	Baritone sax	Bass voice

waveforms, and the relative salience of these esoteric pitch perceptions. The models fall into two groups: those that follow Helmholtz (1875) and attempt to explain the perception of pitch on the basis of the frequency spectra of the sounds, and those that follow Licklider (1951) and emphasize the distribution of time intervals observed in the firing patterns that pitch-producing sounds generate in the auditory nerve. A brief overview of the debate is presented in Sect. 2.4 of this chapter; more extensive discussions are provided in a recent article by Yost (2009) and a recent chapter by de Cheveigné (2005). Despite the passion of the debate between the spectral and temporal modelers, for readers who are simply interested in the relationship between the physics of note production and perception, the pitch of the notes of the main orchestral instruments is simply the psychological correlate of the repetition rate of the waveform that the instrument produces.

With regard to timbre, the instruments of a given family have similar physical shapes, they are made of similar materials, and they are excited in similar ways, so it is not surprising that the instruments of a family produce tones with a similar sound quality, or timbre, that distinguishes the family. The categories of timbre associated with instrument families are labeled with words that describe some physical aspect of the source. So, the trumpet is a *brass* instrument, the clarinet is a *wood-wind* instrument, and the violin is a *string* instrument. The *family* aspect of timbre is largely determined by the shape of the envelope of the magnitude spectrum of the tones that the instrument produces. This aspect of musical perception is also relatively straightforward for sustained-tone instruments.

Within a family of instruments, the different members are distinguished physically by their size, and perceptually by the effect that the size of the instrument's components has on the tones they produce. There are two different aspects to instrument size, and they jointly determine our perception of the *register* of an instrument within its family. In the string family, register distinguishes the violin, viola, cello, and double bass, and the instrument names are normally used to specify the instrument's register. In the string family, as the size of the instrument increases from violin to double bass, the lengths and masses of the strings increase, and so the tones of the larger instruments have lower pitches (on average). The range of pitches that an instrument produces is one of the properties that determine the register we perceive and what instrument we hear within a family. The second aspect of instrument size is the size of the body and it also affects the register we perceive and the instrument we hear; larger bodies go with lower registers. The fact that register depends on two acoustic variables means that the perception of register is somewhat more complicated than the perception of pitch and family timbre. Nevertheless, the principles, as they pertain to the perception of musical tones, are readily comprehensible and they are a prominent topic in this chapter. To begin with, register can be regarded as the perceptual property that enables us to distinguish the size element of instruments within a family (Table 2.1) including the categorization of humans as sopranos, altos, tenors, baritones, or basses. Note that children, when they begin to sing, are sopranos and they progress down in pitch to their eventual range as they grow up.

In summary, the main purpose of this chapter is to describe how the *physical* variables of tone generation are related to the *acoustic* variables of tones as sounds in the air, and how these acoustic variables are related to the *perception* of melodic pitch, family timbre, and register within an instrument family. There is a secondary aspect of register, associated with the perception of individual instruments, that allows us to distinguish the upper and lower notes by the sound of the tones themselves, and we refer to the tones as coming from the upper or lower “register” of a particular instrument, or voice. We return to this secondary aspect of tone perception later in the chapter, once the acoustic properties of sound, and their primary role in perception, have been set out.

2.2 Pulse-Resonance Sounds and Acoustic Scale

The tones that one hears in the natural environment are typically “pulse-resonance” sounds (Patterson et al. 2008), for example, the calls that mammals, birds, frogs, and fish use to declare their territories or attract mates (e.g., Fitch and Reby 2001). The vowels of speech and the sustained tones of orchestral instruments are also pulse-resonance sounds. So they are the normal tones that one hears every day in the man-made environment and in the natural world.

2.2.1 Origin of Pulse-Resonance Sounds

The production of a pulse-resonance sound is conceptually simple. The animal just has to develop some means of producing an acoustic pulse that will, then, resonate in one or more of the structures in the animal’s body. Once the basic mechanism arises in response to the need for communication, evolution can refine the sound with successive modifications to make it more distinctive and efficient. In present-day animals, the pulse generating mechanism typically produces a stream of pulses that occur regularly in time, and in models of tone production, the mechanism that produces the stream of pulses is referred to as “the source” of the sound. The resonances in the animal’s body are collectively referred to as a “the filter,” and in most animals, the filters have evolved to give the animal’s call a distinctive timbre. The stream of pulses with their resonances forms a tone, and these tones provide the basis for animal communication. They also broadcast the species of the caller.

In almost all mammals, the source mechanism is the vocal folds in the larynx at the base of the throat; they produce pulses by momentarily impeding the flow of air from the lungs. The pulses of air then excite resonant cavities in the airway between the larynx and the lips, and this filter of resonant cavities is referred to as the vocal tract. A short segment of a synthetic /a/ that sounds like the vowel in “car” is presented in Fig. 2.1a. The wave shows that the sound is periodic and each cycle contains an acoustic pulse followed by a decaying resonance with a complex shape. A vowel is normally on the order of 100–300 ms in duration, so the complete

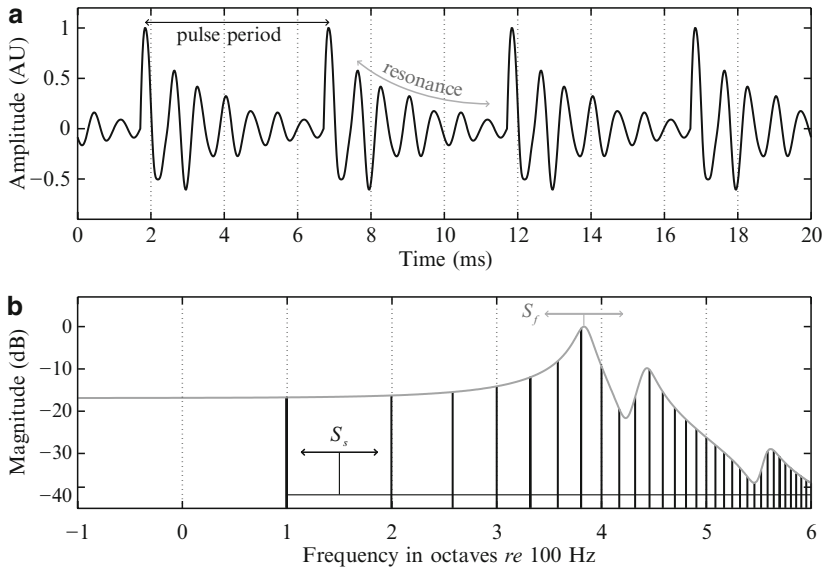


Fig. 2.1 The waveform and magnitude spectrum of a child’s vowel /a/. **(a)** The waveform, which is a plot of acoustic pressure as a function of time, shows a repeating pattern that starts with a pulse. The repetition period, or pulse period, is shown by the black arrow. Each pulse is followed by a resonance that decays in time, as shown by the gray arrow. **(b)** The long-term magnitude spectrum, that is, the distribution of energy across frequency, is composed of harmonics represented by the vertical black lines that form the *fine-structure* of the spectrum. The frequency axis is logarithmic and scaled in number of octaves *re* 100 Hz. The position of the fine-structure, that is, the position of the set of harmonics taken as a unit, is the acoustic scale of the source S_s . This quantity is related to the pulse period shown on the waveform. The spectral envelope, shown in gray, depicts how the resonators in the vocal tract filter the pulses. Its shape determines the vowel type. Its position on the log-frequency axis is the acoustic scale of the filter, S_f

waveform for the /a/ in “car” would contain 20–60 of the pulse-resonance cycles shown in Fig. 2.1a. The waveform repeats every 5 ms so the “repetition rate” of the tone is 200 cycles per second (cps), and this value is used to specify its pitch.

These are the main characteristics of pulse-resonance sounds as they appear in the time domain. Many birds and frogs also excite resonances in their air passages by momentarily interrupting the flow of air from the lungs, although the details of the source and filter mechanisms are somewhat different. Fish do not have air passages but many of them have swim bladders that resonate and function as the filter. The bladder is excited by muscles in the wall of the swim bladder (e.g., the weakfish, *Cynoscion regali*) that produce brief mechanical pulses referred to as “sonic twitches.” This muscle source produces twitches in regularly timed streams (Sprague 2000). A brief introduction to the pulse-resonance sounds produced by animals is presented in Patterson et al. (2008).

Pulse-resonance tones are very different from environmental noises such as wind in the trees or waves on the beach, or man-made noises like extractor fans, jet engines,

or the boiling of a kettle. Noises arise from turbulent systems where the source vibrates randomly. Noise waveforms are not periodic and so they do not produce salient pitch perceptions. The filtering is incidental and evolution is not involved in tuning the filter to make the sound distinctive or improve communication. One continuous noise sounds much like another when they have the same loudness. Perceptually, pulse-resonance tones, with their pronounced pitch and distinctive timbre, tend to capture the listener's attention, whereas continuous noises are commonly ignored.

Returning to musical sounds, the sustained tones that singers produce when the voice is used as an instrument are vowels, and so the singing voice produces pulse-resonance tones. The instruments of the brass, string, and woodwind families also produce pulse-resonance tones (van Dinther and Patterson 2006). Each of the families has a source mechanism that produces regular streams of pulses that are filtered by resonances in the instrument's body (Fletcher and Rossing 1998). Several examples are presented in Sect. 2.3. The remainder of this section describes the acoustic properties of pulse-resonance tones as they appear in the magnitude spectra of the sounds, and how the properties do, *or do not*, vary with the size of the instrument or singer.

2.2.2 Acoustic Properties of Pulse-Resonance Sounds

The set of vertical lines in Fig. 2.1b shows the long-term *magnitude spectrum* of the vowel, that is, the distribution of energy across frequency, averaged over 100 ms, or more, of time. The frequency axis is logarithmic in this case, similar to the place, or "tonotopic," dimension of the cochlea. The vertical lines show that the energy is restricted to frequencies that are integer multiples of a single, *fundamental* frequency, designated F0. The fundamental of this harmonic series, and the frequency spacing between the harmonics (Fig. 2.1b), are the spectral representation of the repetition rate of the sound, which is the inverse of the period observed in the waveform (Fig. 2.1a). In this example, all three of these acoustic variables have the value 200 cps. The dashed line connecting the tops of the harmonics in the lower panel shows the *spectral envelope* of the vowel.

The soft shouldered peaks that appear in the spectral envelopes of speech sounds are referred to as formants. Individual formants are normally designated by the frequency of the peak in the envelope, but the concept of a formant actually includes the shape and width of the envelope in the region of the peak, as well as the peak frequency. The shape that the set of formants collectively impart to the envelope in the spectral domain (Fig. 2.1b), is related to the shape of the damped resonance following each glottal pulse in the time domain (upper panel). The resonators in the bodies of musical instruments do not produce such distinctive formants as the resonances of the vocal tract, but the principles are the same for all pulse-resonance sounds. The shape of the spectral envelope corresponds to the shape of the resonance in the waveform, and this shapes determine the

distinctive sound quality, or timbre, of an instrument family. The set of harmonics that constitute the magnitude spectrum of a sound will be collectively referred to as the *fine-structure* of the spectrum to distinguish the magnitude spectrum (solid vertical lines) from its envelope (gray line).

Now consider the changes that occur in the tones of a specific instrument family as the size of the instrument increases. For example, consider what happens to vowel sounds as children grow into adulthood. When children begin to speak they are about 0.85 m tall, and their height increases by about a factor of two as they mature. In humans (and other animals), the source and filter are components of the body and both the source and the filter increase in size as young mature into adults. With regard to the source in humans, the glottal pulse rate (GPR) decreases by about an octave as the child grows up and the vocal cords become longer and more massive. The decrease in GPR is greater than an octave for males and less than an octave for females, but even for females, it is a large change. With regard to the filter, vocal tract length increases in proportion to height (Fitch and Giedd 1999; Turner et al. 2009, their Fig. 4), and as a result, the formant frequencies of children’s vowels decrease by about an octave as they mature (Lee et al. 1999; Turner et al. 2009). The effects of growth on the fine-structure and envelope of the spectrum of a vowel are quite simple to characterize, provided the spectrum is plotted on a logarithmic frequency scale. In this case, the set of harmonics that define the fine structure of the spectrum (the vertical lines in Fig. 2.1b) moves, *as a unit*, toward the origin as the child matures into an adult. In speech, the pattern of formants that defines a given vowel type remains largely *unchanged* as people grow up (Peterson and Barney 1952; Lee et al. 1999; Turner et al. 2009). In other words, for a given vowel, the shape of the spectral envelope does not change as a child matures; rather, the spectral envelope just shifts slowly toward the origin, moving about an octave in total as a child matures into an adult. Thus, in the current example, the vowel remains an /a/, and does not change to an /e/, an /o/ or an /u/, as a child matures into an adult.

The “position of the spectral envelope of a sound on a logarithmic frequency axis” is a property of a sound as it occurs in the air (Cohen 1993). For a pulse-resonance tone, this property is the *acoustic scale* of the filter that defines the resonances, and in the case of the human voice, it is closely related to vocal tract length (a physical variable). The “position of the fine-structure of the spectrum on a logarithmic frequency axis” is also a property of a sound as it occurs in the air. For a pulse-resonance tone, it is the *acoustic scale* of the source, and in the case of the human voice, it is closely related to glottal pulse rate (a physical variable). The two acoustic scale variables are very useful for summarizing the effects of physical variables such as mass and length on the perceptions produced by instruments, and, as a result, they play a prominent role in the remainder of the chapter. For brevity, “the scale (S) of the source (s)” is designated S_s , and “the scale (S) of the filter (f)” is designated S_f . Turner et al. (2009) have recently reanalyzed several large databases of spoken vowels and shown that almost all of the variability in formant frequency data that is not vowel-type information is S_f information. To reduce confusion between the two acoustic scale variables, S_s and S_f , we use cycles per

second (cps) for the units of the scale of the source, S_s , and kiloHertz (kHz) for the scale of the filter, S_f , since the position of the spectral envelope and the unit for the frequency dimension of the magnitude spectrum is Hertz.

In summary, the important distinctions for the remainder of the chapter are as follows:

1. The pulse rate of the source is a physical variable (e.g., GPR). It determines the repetition rate of the wave, which is known as the acoustic scale of the source, S_s . Repetition rate and S_s are both acoustic variables, and they in turn, determine the pitch of a pulse-resonance tone. Pitch is a perceptual variable.
2. The size of a resonator in the body of a person or an instrument is a physical variable (such as length or volume). It determines the rate at which the resonance oscillates in the waveform (van Dinther and Patterson 2006), and it determines the *position* of the spectral envelope along the frequency axis of the magnitude spectrum. It is known as the acoustic scale of the filter, S_f , and it is an acoustic variable that affects the perception of source size and the perception of register within an instrument family.
3. The shape of the spectral envelope determines the instrument family aspect of timbre.
4. Register is the term used to describe the joint action of the acoustic variables, S_s and S_f , on the perception of musical tones and instruments. The values of S_s and S_f reflect the physical sizes of the source and filter in the instrument, respectively, and so the perception of register is closely related to the perception of instrument size, or singer size. The vocal terms *soprano*, *alto*, *tenor*, and *bass* are commonly used to specify register within families, as in tenor sax or bass fiddle.
5. Finally, note that the voice differs from other instruments with respect to timbre, in one important regard. When vowel type changes, say, from /a/ to /i/, the shape of the envelope changes. The shape does not change with the size of the singer from child to adult, whereas the acoustic scale values, S_s and S_f , do. So, different vowels are like different instrument families in the perception of musical tones. One useful, and reasonable, way to think of vowels is that they form a cluster of instrument families (unified by the fact that they are perceived to come from humans) and that the differing timbres of the members of this family are somehow more similar to each other than they are to the timbres of other musical instrument families.

2.2.3 Terminology

2.2.3.1 Source

There are many meanings of the word *source* in the description of sounds and how they are produced. To avoid confusion when reading this chapter, focus on what the *source* is a source *of*. So, when listening to an orchestra, one specific musician, and

the instrument they are playing, jointly form the source of one of the streams of musical tones that the orchestra is producing. In contrast, the “source” of the *energy* in these tones is the arm of the musician, in the case of string instruments, and the diaphragm of the singer in the case of a vocalist. The “source” in a *source-filter system* is a mechanism in between the source of the energy and the complete instrument in combination with the musician. In the source-filter description of tone production, the word “source” means the mechanism that produces the stream of abrupt amplitude changes, or pulses, which subsequently excite the set of resonances in the body of the instrument, or the vocal tract of the singer. It is a specialized meaning of the word “source,” but it is straightforward and it is the only use of the word “source” in this chapter.

2.2.3.2 Noise

Throughout the current chapter, we use the word “noise” as an acoustic term that refers to the fact that the waveform is aperiodic and the amplitude varies randomly with time. These sounds are typically heard as background sounds and do not draw your attention. There is, of course, another use of the word “noise” that can occur in a musical context. For example, when there are competing sounds in an environment, perhaps a Mozart symphony on the radio and a rock concert on television, an individual listener might say, “Turn off that noise!,” referring to the source which is interfering with the source they are trying to hear. The current chapter is not concerned with multisource environments and so the latter use of “noise” does not arise in this chapter.

2.2.3.3 Scale

In the phrase “acoustic scale” the word *scale* is being used in the mathematical sense, rather than the musical sense. In mathematics “a scale factor” is a number that tells you how big one value is relative to another. A musical scale is a set of frequency intervals within an octave. There is a connection between the two uses of scale inasmuch as the intervals of a musical scale (such as a fifth) are defined by specific scale factors (~ 1.5 in the case of a fifth), but acoustic scale refers to a single value rather than a set of musical scale values.

2.3 The Pulse-Resonance Tones of Musical Instruments

This section describes how the sustained-tone instruments of the orchestra produce their tones, and the relationship between the physical properties of the instrument on the one hand, and the three main acoustic properties of these sounds on the other hand.

2.3.1 *The Source of Excitation and the Acoustic Scale Variable, S_s*

In general terms, the “source” in these instruments is a highly nonlinear, resonant system that produces a temporally regular stream of acoustic pulses. The mechanism is conceptually similar for the voice, brass instruments, and woodwind instruments; in these instruments, the source momentarily interrupts the flow of air from the lungs, and it does so regularly in time. The individual mechanisms are, however, quite diverse. For example, the source is the vocal folds in the case of the voice; whereas, in brass instruments, it is the lips coupled to the main tube via the mouth-piece; and in the woodwinds, it is the lips coupled to the main tube via the reed. In string instruments, the mechanism is completely different; it is the bow coupled to a string. Despite the diversity of mechanisms, all of the sources produce streams of very precise acoustic pulses (brass and woodwinds), or abrupt changes in amplitude (strings) that function in a similar way. As a result, the sound waves produced by sustained-tone instruments are all pulse-resonance sounds. (In Fourier terms, the overtones of the pulse rate are locked to the pulse times both in frequency *and phase* up to fairly high harmonic numbers.)

The acoustic scale of the source of excitation is termed the *source scale* or S_s ; it is effectively the repetition rate of the wave as it occurs in the air between the instrument and the listener. S_s is determined by physical properties of the instrument, like length and mass, which are not themselves acoustic variables. S_s largely determines the pitch we hear, but S_s is not itself an auditory variable. It is an intervening, acoustic variable that describes a property of the sound in the air, and it should be distinguished from pitch, which is the auditory variable of perception. The relationship between S_s and the physical variables of the instrument will be illustrated by comparing how S_s is determined in the vocal tract and in string instruments.

2.3.1.1 **The Source of Excitation in the Human Voice**

The vocal folds produce glottal pulses in bursts and, although the vocal folds are rather complicated structures, the effects of the physical variables on the rate of pulses can be described using the expression for a tense string. The glottal pulse rate, GPR, is largely determined by the length, L , mass, M , and tension, T , of the vocal folds, and the form of the relationship is

$$\text{GPR} \propto \sqrt{\frac{T}{ML}} \quad (2.1)$$

Two of these physical variables are determined by the size of the person – the length and mass of the vocal folds. Both of these variables increase as a child

grows up, and both of these terms are in the *denominator* on the right-hand side of the equation, so as the child *increases* in height the pitch of the voice *decreases*. The average GPR for small children is about 260 cps, both for males and females. For females, GPR just decreases with height throughout life dropping to, on average, about 160 cps in adult women. For males, GPR decreases with height until puberty at which point the vocal folds suddenly increase in mass and the GPR drops to, on average, about 120 cps. So the length and mass of the vocal folds are a major determinant of vocal register, that is, whether a singer is a soprano, alto, tenor, baritone, or bass.

To produce a melody, a singer varies the tension of his or her vocal folds. So learning to sing in tune is largely a matter of learning to control the tension of the vocal folds – holding the tension fixed during sustained notes and changing it abruptly between notes. Tension is in the *numerator* of the mathematical expression (2.1), and so as a singer *increases* the tension, he or she *increases* the GPR. There is considerable overlap in the note ranges of the soprano, alto, tenor, baritone, and bass voices; in fact, the highest note of a bass is typically a note or two above the lowest note of a soprano. The effect of all three of these variables (T , M , and L) on GPR is constrained by the fact that the GPR value is related to the square root of these variables. So, for example, a singer has to change the tension of the voice by a factor of four to produce a one-octave change that would double the GPR.

In summary, for a specific individual, the size of the vocal folds (length and mass) determines the individual's long-term average GPR, and it determines the S_s component of the register of their voice. The tension of the vocal folds is varied to produce a melody. So, the long-term average S_s value, calculated over a sequence of musical phrases, reveals the register of the singer's voice; short-term deviations of S_s from the longer-term average, in discrete steps with regular timing, are the hallmarks of vocal melody.

2.3.1.2 The Source of Excitation in the String Family

The excitation mechanism in stringed instruments is the string pushed by the bow. As the musician draws the bow across a string, the string is pushed or pulled away from its resting position until the tension becomes too great, at which point, it snaps back, producing an abrupt, unidirectional change in amplitude. The direction is opposite to the direction that the bow is moving. The result is, nevertheless, a pulse-resonance sound inasmuch as the harmonics are locked in phase, and the internal representation of the sound has a pulse-resonance form in any given frequency band. Although the bow-string system is rather complicated physically (McIntyre et al. 1983), the relationship between pulse rate, PR, and the main physical variables is the same as for the vocal folds, namely,

$$\text{PR} \propto \sqrt{\frac{T}{ML}} \quad (2.2)$$

In this case, however, T , M , and L refer to the tension, mass, and length of the string, rather than to the corresponding properties of the vocal folds. The two physical variables associated with the size of the source (the length and mass of the string) are the most important excitation variables in this family of instruments and they each have *two* roles to play. Consider first the pulse rates of the open-strings on these instruments: Both the mass and length variables are in the denominator on the right-hand side of the equation, so *increases* in size, be they length or mass, lead to *decreases* in pulse rate. For a given member of the family (violin, viola, cello, or contra bass), the length of the four strings is fixed, and as the size of a family member increases, the string length gets longer in discrete steps. As a result, string length plays an important role in determining register within the string family. The mass of the string increases with its length, so it also contributes to the register we perceive. Mass also plays an important role in determining the range of notes that an individual instrument can play; the mass is varied across the four strings to extend the range beyond that which can be provided on any one string. Finally, the musician varies the length of individual strings to produce the different tones within that string's range.

Instrument makers are very adept at using mass and length to vary the pulse rate of notes within a family. If a musician depresses the lightest string on the largest instrument (the contra bass) at a point near the bridge on the neck, the pulse rate of the note will actually be a little higher than the pulse rate of the open-string note of the heaviest string on the smallest member of the family (the violin). In both cases, the notes are just below middle C on the keyboard.

2.3.1.3 Excitation Mechanisms of the Woodwind and Brass Instrument Families

The excitation of woodwind and brass instruments is described in terms of fluid mechanical “valves” that momentarily close the flow of air through the instrument. The closure causes a sharp acoustic pulse that resonates in the tube beyond the mouthpiece. For woodwind instruments, the valve is the reed in conjunction with the lips. For brass instruments, the source is not clearly localized within the instrument. The source of *energy* is the stream of air produced by the player who controls the pressure with the tension of the lips. The source of excitation is pulsatile because the mouthpiece is coupled to the tube between the mouthpiece and the bell (i.e., the body of the instrument), and the tube can only resonate at certain frequencies. Thus, the pulses originate from the lips, but the pulse rate is determined by the effective length of the tube, and this functional tube length is varied by the valves (or the slide) to control the pulse rate of the note.

Despite the complexities of excitation, these two families of instruments produce pulse-resonance sounds in which the acoustic scale of the source S_s controls the repetition rate of the note, and thus contributes to define the instrument's register within its family. The pulsatile nature of the excitation generated by these systems, and the temporal regularity of the pulse stream, mean that the dominant components

of the spectrum are strictly harmonic and they are phase locked (Fletcher and Rossing 1998). Fletcher (1978) provides a mathematical basis for understanding the origin of the phase locking, which is referred to as mode locking in musical instrument theory. Detailed descriptions of the mechanisms are provided in Benade (1976), Fletcher (1978), and McIntyre et al. (1983); a brief overview is provided in van Dinther and Patterson (2006).

2.3.1.4 Summary of the Role of S_s in Determining Melody and Register Within a Family

Comparison of the excitation mechanisms for the different instrument families shows that these mechanisms are similar, inasmuch as they all produce regular streams of pulses and the pulse rate is affected in the same way by the size of the components in the source. As a result, pulse rate decreases as instrument size increases in all of these instrument families. At the same time, the method whereby the pulse rate is varied to produce a melody is fundamentally different: the variable that controls pulse rate in the voice is the tension of the vocal folds, and the singer *increases* the tension to *increase* the pulse rate; whereas the variable that controls pulse rate in string instruments is string length, and the musician *decreases* the length to *increase* the pulse rate. The brass and woodwind instruments are like the strings, inasmuch as the pulse rate is varied to produce a melody by varying the length of part of the instrument; brass and woodwind instruments are different from the strings inasmuch as the length in this case is tube length rather than string length.

Although different instrument families employ very different mechanisms to produce acoustic pulses (and it is important for musicians to understand something of these mechanisms in order to play their instruments properly), all of these instruments nevertheless produce pulse-resonance tones, and the melody information in music is a sequence of pulse-rate values that specify the momentary acoustic scale of the source of excitation. Although the relationship between the physical variables involved in instrument excitation and the repetition rate of a given note is complex, the relationship between the acoustic-scale variable, S_s , which summarizes the action of the source, and the pitch we perceive is straightforward.

2.3.2 *The Filtering of the Excitation Pulses and the Acoustic Scale of the Filter, S_f*

The “filter” in musical instruments is a set of resonators that increase in size with register within an instrument family, and together the resonators determine the acoustic scale of the filter, S_f . Each of the pulses produced by the excitation mechanism of a sustained-tone instrument is filtered by body resonances within the instrument. In the time domain, it is these resonators in the body of the instrument that produce the resonances that appear attached to each pulse in the waveform (e.g., Fig. 2.1a).

In the frequency domain (e.g., Fig. 2.1b), the body resonances produce the distinctive shape of the envelope of the magnitude spectrum, and consequently, they determine the timbre of the family. In the case of the voice, the dominant resonances are associated with the larger cavities of the vocal tract (Chiba and Kajiyama 1941; Fant 1960). The tongue makes a constriction in the vocal tract that divides it into a mouth cavity and a throat cavity. These cavities resonate like tubes and/or bottles and they introduce formant peaks into the vowel spectrum (Fig. 2.1b). The tongue position is varied to produce the different vowels. This changes the relative sizes of the cavities, and thus, the relative positions of the formants in the spectrum (Chiba and Kajiyama 1941; Fant 1960). For stringed instruments, the most important resonances are associated with the plates of the body (wood resonances), the body cavities (air resonances), and the bridge (structural resonances) (Benade 1976). For brass and woodwind instruments, the prominent resonances are associated with the shape of the mouthpiece, which acts like a Helmholtz resonator, and the shape of the bell, which determines the efficiency with which the spectral components radiate into the air (Benade and Lutgen 1988). Woodwind instruments are like brass instruments, but the materials are different. So, just as there are many source mechanisms for generating the pulse stream, there are many systems of body resonances that lead in turn to many distinctive spectral envelopes.

Within a family of instruments, the most prominent distinction between the members of the family is the size of the body of the instrument, and the primary effect of instrument size on the perception of register within a family is straightforward (van Dinther and Patterson 2006): If the size of an instrument is changed while keeping its shape the same, the result is a proportionate change in S_f , the acoustic scale of the filter mechanism in the body of the instrument. That is, if the three spatial dimensions of an instrument are increased by a factor, a , keeping the materials of the instrument the same, the natural resonances decrease in frequency by a factor of $1/a$. The shape of the spectral envelope is preserved under this transformation, and so, if the spectral envelope is plotted on a log-frequency axis, the envelope shifts as a unit toward the origin, without changing shape, and the change in S_f will be the logarithm of the relative size of the two instruments: $\log(1/a)$. This uniform scaling relationship is called “the general law of similarity of acoustic systems” (Fletcher and Rossing 1998), and it is used to produce much of the difference in S_f between the tones produced by different instruments within a family. Numerical examples illustrating how the spatial dimensions of an instrument affect its resonances are provided by van Dinther and Patterson (2006).

Comparison of the filter systems of the different instrument families shows that the spectral envelope is affected in the same way by changes in the size of the filter-system components; specifically, the resonant frequencies *decrease* as body size *increases* and so the spectral envelope shifts toward the origin as the sizes of the components increase. So size affects the filter system in the same way as it affects the excitation mechanism. It is another example of the fact that bigger things vibrate more slowly. The wood-plate and bridge resonances of the string-family filter system are complex, and they are fundamentally different from the bell and mouthpiece resonances of the brass-family filter system, which are also complex.

Despite the complexity of the relationship between the physical variables involved in body filtering and the shape of the resultant spectral envelope, the relationship between the acoustic properties and the perception of the notes is fairly straightforward. The shape of the spectral envelope determines the family aspect of timbre; the acoustic scale of the filter, S_f , determines the register we perceive, and thus, which instrument within the family. In all of these instrument families, the register decreases from soprano to bass as instrument size increases and the spectral envelope shifts toward the origin.

2.3.3 *Constraints on the Acoustic-Scale Variables in Orchestral Instruments*

In Sects. 2.3.1 and 2.3.2, the relationship between the physical variables involved in the production of musical tones, and the acoustic scale of the source, S_s , and the filter, S_f , was presented in theoretical terms without reference to the practicalities of constructing and playing instruments. In the real world, it turns out that it is not possible to simply scale the spatial dimensions of instruments to achieve registers ranging from soprano to bass in most instrument families; the bass member would be too large and/or the soprano member too small. This section reviews the spatial scaling problem, and describes how the instrument makers produce tones with a wide range of acoustic scale values without using excessively large or small instruments.

The spatial scaling problem arises from the desire to simultaneously satisfy three design criteria for families of sustained-tone instruments: The first criterion is that the instruments should produce notes that are heard to have a strong musical pitch, whose clarity and salience provide for effortless communication of melodies and their variations. This places an important constraint on the relationship between the acoustic scale variables, S_s and S_f . The instrument's filter system must resonate at frequencies corresponding to the first 10 harmonics of the pulse rate of each note that the instrument is intended to play; that is, the instrument must emit significant amounts of acoustic energy in the range from the pulse rate of each note to three octaves above that pulse rate. This is necessary because the pitch of notes where the energy is carried by harmonics above about the tenth is not sufficiently salient to support accurate perception of novel melodies (Krumbholz et al. 2000; Pressnitzer et al. 2001). The second criterion is that the members of each instrument family should, together, produce notes that cover a significant portion of the musical scale, which for the keyboard encompasses about seven octaves from, say, 27.5–3520 cps. When combined with the first criterion, the second criterion effectively requires that the instruments of a given family have matched S_s and S_f values for all of the registers in the range from soprano to bass. This is a very demanding constraint, particularly when combined with the third criterion, which is that the instruments should be playable and portable. This last, practical constraint places limitations on the sizes of instruments which, in turn, means that the desired range of notes cannot be achieved by simply scaling instrument size in accordance with the law of acoustic similarity.

There are problems for the instrument maker at both ends of the register range. For example, in the string family, there is a limit to how short the neck can be on the smallest member of the family (the violin) if the contact points where the string is pressed onto the neck are to be far enough apart for a musician to play the notes of a melody accurately and quickly. And at the other end of the range, if the instrument maker attempts to scale up the soprano version of the family to provide the bass member, the instruments become too large to play *and* too large to carry. Hutchins (1967, 1980) described the problems encountered when you try to construct a family of eight stringed instruments covering the entire range of orchestral registers based on the properties of the violin. The double bass member of the family would have to be six times the size of the violin, if simple scaling of instrument dimensions were to be used to provide a shift of six octaves in the spectral envelope. The length of a violin is about 0.6 m, so the double bass in this hypothetical family would have to be 3.6 m tall. The lower notes on the strings of such a double bass would not be reachable for most musicians and the instrument would not be portable. So, the problem is this: Although instrument makers can scale the dimensions of instruments to achieve much of the desired change in S_s and S_f , it is not possible to use the scaling of spatial dimensions, on its own, to provide the full range of registers in each family, and at the same time, ensure that the pitch of each note is sufficiently strong to support accurate melody perception.

So how do instrument makers solve this problem, and how do they construct families of instruments that produce tones with salient pitches over the full range of registers from soprano to bass – instruments that are, at the same time, playable and portable? The first criterion of instrument production is immutable; the instrument must produce energy in the first three octaves of the pulse rate if the note is to have a well defined pitch. The third criterion is essential; the instruments have to be playable and portable. So how do the instrument makers provide a wide range of notes on instruments with manageable sizes? This is where the knowledge and craft of the instrument maker come to the fore. What is required is not that the soprano instruments be excessively small and the bass instruments be excessively large; what matters is that the instruments produce tones with a wide range of S_s and S_f values, and that the S_s and S_f values are coordinated throughout the range. So what the instrument makers have done is find ways of extending the range of S_s and S_f values beyond what is practical with spatial-dimension scaling, by adjusting other physical properties of the instruments such as the mass of the strings, the thickness of the plates or the depth of the volume of the air cavity. They scale the physical dimensions of the family so that the largest member is portable and the smallest member is playable, and then they adjust other physical properties of the instrument to achieve the desired acoustic scale values for the source mechanism and the filter system (e.g., Schelleng 1963).

Consider the case of the source scale in the string family: The strings on the larger members such as the cello and contra bass are not as long as the law of acoustic similarity would require because it would make the instruments unwieldy. The instrument makers increase the linear mass of the strings (the mass per meter)

by winding metal coils around the string. This increased mass causes the strings to vibrate more slowly as illustrated by Eq. (2.2). The instrument makers use a change in mass to obtain the lower ranges of notes on the lower strings of any given member of the family.

With regard to the filter scale in the string family: The filter systems of the larger members of the family are not as large as the law of acoustic similarity would require, because it would make the instruments too heavy and too large. The instrument makers adapt the characteristics of the instruments to preserve the sound quality while making them usable at the same time. The main resonance is driven by the cavity mode of the body which functions like a Helmholtz resonator. The volume of the instrument as well as the surface area of the f -holes are the key parameters. The open strings of the cello are tuned to pulse rates three times lower than those of the violin. However, the plates of the cello's body are only 2.1 times larger than those of the violin (Schelleng 1963), while the rib height of the cello is about four times that of the violin (Fletcher and Rossing 1998). Thus the volume of the cello is 17 times larger than that of the violin; this is equivalent to uniform spatial scaling by a factor of 2.6. To lower the body resonances to the desired values, the instrument makers vary the mass, thickness and arching of the body plates. Specifically, the body plate of the cello is made proportionally thinner than that of the violin which lowers the body resonance frequency (e.g., Molin et al. 1988).

Having established that the acoustic scale variables are balanced in the sustained-tone instruments of the orchestra, we can return to the secondary aspect of register, associated with the perception of tones from a single instrument, that is, the within-instrument register. Register, in this sense, is “a part of an [instrument's range] having a distinctive tonal quality” (Kennedy 1985, p. 585). So we speak of the chest and head registers of an individual's voice, or the upper and lower register of an instrument's range. In acoustic scale terms, the perception of register within an instrument's range, is a perceptual distinction concerning the relative values of S_s and S_f . When the S_s values of a succession of notes are high relative to the S_f of the singer or the instrument, we perceive that the person is singing, or the instrument is playing, in the upper register, and vice versa.

Finally, note that the range of tones covered by the registers of the voice, from soprano to bass, is only about four octaves in total (from about C6 down to slightly over C2). The range of the string-family instruments (taken together) covers almost seven octaves (from just under C8 to just over C1). The singing teacher can help a vocalist strengthen tones toward the ends of their natural range, but they cannot stretch the vocal tract length or add significant mass to the vocal folds.

In summary:

1. Although the physics of the source mechanisms that excite the sustained-tone instruments is complicated, and the physics varies markedly from family to family, the acoustic scale of the source, S_s , provides a convenient summary of the action of the source as it pertains to tone perception. The source determines the repetition rate of the wave, or the position of the fine structure of the magnitude spectrum (on

- a log frequency axis), and this, in turn, determines the pitch of the tone, and contributes to the perception of an instrument's register within its family.
2. Although the physics of the resonance mechanisms that filter the source waves is complicated, and the physics varies markedly from family to family, the acoustic scale of the filter, S_f , provides a convenient summary of the action of the filter with regard to its contribution to the perception of an instrument's register within its family.
 3. Within a family, when source size is increased to increase the acoustic scale of the tones and lower the pitch, the acoustic scale of the filter has to be increased to maintain the distinctive timbre of the family, and to ensure that the tones continue to produce a strong pitch. At the same time, the increase in filter scale contributes to the lowering of the perception of the register of the instrument within its family.
 4. Within a family, it is not possible to produce tones whose pitches span the entire range of the keyboard simply by varying the spatial dimensions of the source and the filter. To achieve the desired acoustic scale values, and the appropriate balance between the acoustic scale values, the instrument maker has to vary other physical properties like the mass of the strings and the stiffness of the plates.

2.4 The Auditory Representation of Pulse-Resonance Sounds and Acoustic Scale

This section presents a brief description of a time-domain model of auditory perception to show how the auditory system constructs our internal representation of musical tones, and to illustrate how the acoustic scale variables appear in this representation of sound. The internal representation is referred to as an auditory image and the stages of the auditory model are intended to simulate all of the auditory processing required to transform a sound into our initial perception of that sound (Patterson et al. 1992, 1995). The processes are analogous to those that the visual system uses to convert light entering the eye into an initial visual image of that light. Although the algorithms used to simulate the construction of the auditory image are straightforward in signal processing terms, auditory models are not commonly used to explain the perception of tones in music and speech research. The most common representation of sound in these research communities is the spectrogram, which is a temporally ordered sequence of magnitude spectra. The spectrogram is a linear-time, *linear*-frequency representation of sound, and it is normally plotted with time on the abscissa (x -axis) and frequency on the ordinate (y -axis) so that time progresses from left to right as the sound progresses. This section begins with a comparison of two auditory images (shown in Fig. 2.2) that illustrate the essentials of the auditory image as it pertains to the perception of musical tones, and how this representation of sound differs from the spectrogram.

2.4.1 Auditory Images

There are now a number of time-domain models of auditory processing that attempt to simulate the neural response to complex sounds like musical notes at a succession of stages in the auditory pathway, and which produce representations of sound that might be regarded as auditory images (e.g., Slaney and Lyon 1990; Meddis and Hewitt 1991; see de Cheveigné 2005 for a review). In these models, the auditory image is typically constructed in four stages that respectively simulate the operation of (1) the outer and middle ear, (2) the basilar partition, (3) the inner hair cells along the basilar partition, and (4) the temporal integration mechanism in the mid-brain. The Auditory Image Model (AIM; Patterson et al. 1992, 1995) is used to illustrate the construction of auditory images and the form of acoustic scale information in the image, as we currently understand it. What differs from one time-domain model to another is the degree to which they attempt to simulate the details of auditory processing in each stage, and the theoretical bases for the mechanisms chosen to represent these auditory processes. The differences are not particularly important for present purposes because the section is just intended to illustrate the general form of the internal representation of sound and the form of the acoustic scale variables in the internal representation.

The auditory image of a baritone singing the vowel /a/ on the note G2 is presented in Fig. 2.2a, and for comparison, the auditory image of a French horn playing the same note is shown in Fig. 2.2b; the figure is reproduced from van Dinther and Patterson (2006), which provides a more detailed description of the image construction process. The auditory images are the large, two-dimensional, “waterfall” plots; the dimensions of the auditory image are time-interval on the abscissa (from 1 to 35 ms *increasing* toward the left) and frequency on the ordinate (from 0.1 to 6.0 kHz). The properties of the auditory image are introduced with

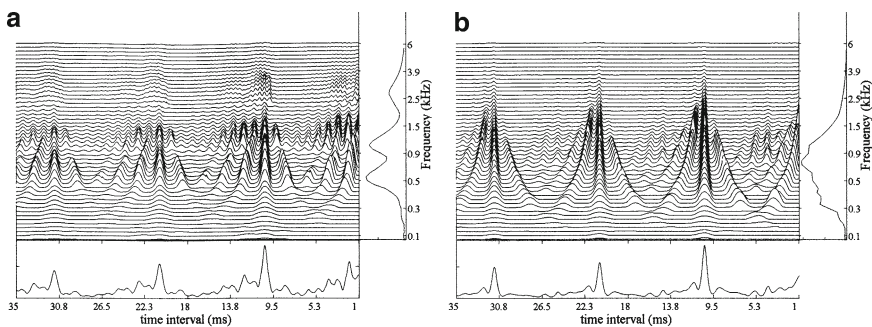


Fig. 2.2 Auditory images of the note G2 (198 cps) as sung by a baritone (a) and as played by a French horn (b). The waterfall plot represents the strobe-stabilized neural activity as a function of time interval since the last strobe time in each frequency channel (see text and Fig. 2.3 for details). The lower profile on each panel is the summary temporal profile. The peaks in this profile show the repetition rate of the sound. The height of the peaks relative to the baseline represents pitch strength (From van Dinther and Patterson 2006)

reference to the four stages of processing used to construct them, and the aspect, or aspects of the auditory image that each stage of processing imparts to the image. The vertical profiles to the right of each image, and the horizontal profile below each image, are introduced once the description of the auditory image itself is complete. (There are multipanel figures in van Dinther and Patterson (2006) that show how the auditory images in Fig. 2.2 were constructed, and how they change as the acoustic scale of the source and the acoustic scale of the filter vary.)

The first stage of processing simulates the effect of the outer and middle ear on incoming sound as it travels from the air through to the cochlea. It is these structures that determine the lower and upper frequency limits for human hearing in young normal listeners. Accordingly, from the perspective of music perception, the first stage determines the range of frequencies that young people normally hear, which is from about 0.1–12 kHz. The vertical dimension of the auditory image is the frequency dimension and so the first stage of processing determines the upper and lower bounds of the auditory image and how activity dies away as it approaches the edges of the image. In AIM, the weighting function is based on the loudness model of Glasberg and Moore (2002). In the case of speech and music, there is very little energy in the region above about 6 kHz, and what is there has very little effect on our perception of musical tones and speech sounds, so the plot of the auditory image is normally limited to 6 kHz as in the images presented in Fig. 2.2.

The second stage of processing simulates the spectral analysis performed in the cochlea by the basilar membrane in conjunction with the outer hair cells and the tectorial membrane; these structures are collectively referred to as the “basilar partition.” The spectral analysis creates the tonotopic dimension along the basilar partition, and it creates the acoustic frequency dimension of auditory perception shown as the vertical dimension in the auditory image. In AIM, as in most time-domain models of perception, the spectral analysis is simulated with a bank of “auditory filters.” Each filter creates a “frequency channel” in the auditory image; that is, the filter passes acoustic energy in a small frequency region about its “center frequency,” and outside this “pass-band,” the filter progressively attenuates acoustic energy as the frequency of that energy diverges from the centre frequency of the filter. This is the essence of an auditory filter. The width of the pass-band of the auditory filter increases with its centre frequency, and the spacing of the filters along the frequency dimension increases with center frequency. As a result, the tonotopic dimension of the cochlea is a quasi-logarithmic frequency axis as shown in the auditory images of Fig. 2.2. In the current version of AIM, the auditory filter is the compressive, gammachirp auditory filter (Irino and Patterson 2001; Patterson et al. 2003).

Each of the lines in the auditory image shows the recent history of activity in a specific frequency channel; the vertical position of the low-level activity in the channel shows the center frequency of each filter. The activity in adjacent channels is correlated and, as a result, the set of filter outputs gives the visual impression of a surface in auditory image space. The surface is AIM’s simulation of the internal representation of sound that is assumed to be the basis of one’s initial perception of a sound. The tones of music produce distinctive structures in the auditory image as illustrated in Fig. 2.2; the structures are referred to as “auditory figures” because

they stand out like figures when presented in background noise (Patterson et al. 1992). The tonotopic dimension of the auditory image is similar to the frequency dimension in Fig. 2.1b insofar as it is quasi-logarithmic; it differs from the strictly logarithmic frequency dimension of Fig. 2.1b inasmuch as the density of channels decreases somewhat below about 0.5 kHz (e.g., Moore and Glasberg 1983).

In the current version of AIM, the auditory filter is the *compressive, gammachirp* auditory filter (Irino and Patterson 2001; Patterson et al. 2003). For readers with an interest in the details, the gammachirp auditory filter is a development of the gammatone auditory filter (Patterson et al. 1995; Unoki et al. 2006). The gammatone filter is essentially symmetric and it is linear, that is, it does not change shape with stimulus level. The gammachirp auditory filter is asymmetric and the asymmetry varies with stimulus level, as dictated by human masking data (Unoki et al. 2006). In the dynamic version of this gammachirp filter (Irino and Patterson 2006), a form of fast-acting compression is incorporated into the auditory filter itself. The compression responds to level changes *within* the individual cycles of pulse-resonance sounds and, as a result, the filter restricts the amplitude of the pulse and amplifies the resonance relative to the pulse in each cycle (see Irino and Patterson 2006, their Figs. 7 and 9).

The third stage of processing simulates neural transduction, that is, the conversion of basilar partition motion into neural activity in the cochlea at the input to the auditory nerve. In AIM, neural transduction is assumed to take place separately in each frequency channel. Specifically, the amplitude versus time wave that flows out of each auditory filter is (1) half-wave rectified (that is, the negative values are set to 0) and (2) low-pass filtered to simulate the upper limit on the firing rate of auditory nerve fibers. The result is a simulation of the aggregate firing of all of the primary auditory nerve fibers associated with that region of the basilar membrane (Patterson 1994a); this function is referred to as a neural activity pattern (NAP). The rapidly oscillating function in Fig. 2.3 shows the NAP flowing from a single auditory filter in response to an /a/ vowel with a GPR of 116 cps and a period of 8.6 ms. The auditory filter is centered just above 1.0 kHz, so the individual cycles of the NAP are just under 1 ms in duration. Each cycle of the vowel produces a distinct cycle of activity in the NAP. There is one of these NAP functions for each of the filters in the filterbank, and together they simulate the response of the cochlea to the vowel.

The fourth stage simulates auditory temporal integration and it converts the set of NAP functions flowing from the auditory filterbank into AIM's simulation of our auditory image of a sound, that is, the neural representation that forms the basis of what we perceive when presented with a sound. In auditory models, this fourth stage of processing is currently hypothetical, in the sense that we do not know precisely how or where it is performed. The reason why perceptual models require a fourth stage is because the time scale of level variation in the NAP functions is not compatible with our perception of sounds; it is clear that there must be some form of temporal integration in the system prior to the neural representation that is the basis of perception. Consider the NAP function in Fig. 2.3: It shows the response to a little over three cycles of the vowel (a total duration of only 0.03 s), so a 1-s segment of the vowel with 116 cycles would be about 30 times the length of the

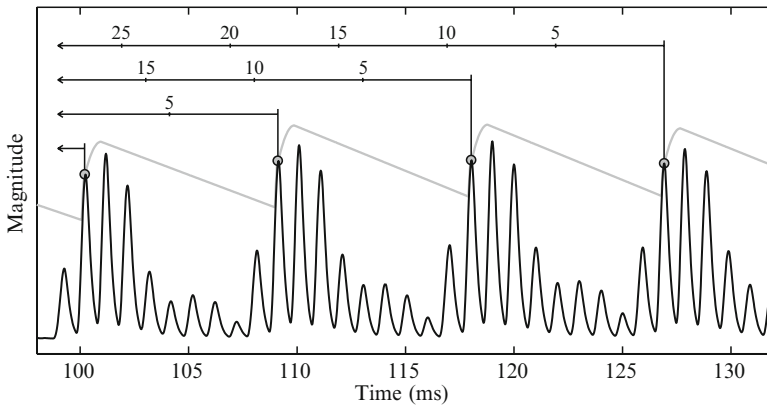


Fig. 2.3 Detail of the neural activity pattern produced by an /a/ vowel at the output of the auditory filter centered at 1018 Hz. The gray line shows the adaptive threshold used to calculate the grey dots, which show the strobe points. The vertical lines and backward arrows show how time intervals are calculated from each of the strobe points backwards in time to earlier points in the pattern, and generate the NAP segment that is added into the corresponding channel of the auditory image to produce the stabilized version presented in Fig. 2.2

segment shown in Fig. 2.3. If Fig. 2.3 were a real-time display (like the neural representation that we perceive), these 30 cycles of the NAP would flow very rapidly from right to left across the display in the course of 1 s, and it would just be a blur. So if the NAP functions were the basis of perception, we would not be able to use the fine-grain temporal information in the NAP functions. However, perceptual research on pitch and timbre indicates that at least some of the fine-grain, time-interval information in the NAP functions is preserved in the auditory image (e.g., Patterson 1994a, b; Yost et al. 1998; Krumbholz et al. 2003). This means that the temporal integration process used to construct the auditory image cannot be simulated by a running temporal average process, like that used to construct the spectrogram because an averaging process would blur the temporal fine structure within the averaging window (Patterson et al. 1992, 1995).

Patterson et al. (1992) argued that it is the fine-structure of periodic sounds that is preserved rather than the fine-structure of aperiodic sounds (e.g., noises), and they showed that the fine-structure of periodic sounds could be preserved by a form of “strobbed temporal integration” controlled by an adaptive threshold. The adaptive threshold for the vowel NAP in Fig. 2.3 is shown by the line with gray dots above the NAP function. It is a form of temporal envelope which emphasizes where the individual cycles of the NAP function start (the dots). These strobe points are used to direct the temporal integration process as indicated by the vertical lines and horizontal arrows above each strobe point. As the start of each new cycle of the NAP function is identified (the dots), a section of the NAP function from the strobe point back to 35 ms before the strobe point (the horizontal lines), is copied and added as a unit into the corresponding channel of the auditory image. In the process the strobe *time* in the NAP function is subtracted from absolute time in the NAP and so, in the

auditory image, the activity associated with any given strobe extends from 0 ms in the auditory image (Fig. 2.2), backwards for 35 ms. As the activity in successive cycles is very similar for pulse-resonance sounds, successive cycles sum to produce a stabilized representation of the pattern in the NAP.

The set of all image channels (one for each filterbank channel) is AIM's representation of our internal auditory image, and the auditory images in Fig. 2.2 were constructed in this way (Patterson et al. 1992, 1995; Patterson 1994a, b). The image decays fairly slowly with respect to the rate of cycles in pulse-resonance sounds (specifically with a half life of 30 ms). So a stabilized version of the neural pattern within the cycle of the sound builds up in the auditory image when the sound comes on and stays there as long as the sound is stationary. When the sound goes off, it decays away to nothing in about 100 ms.

More detailed descriptions of auditory image construction are presented in Patterson et al. (1995), van Dinther and Patterson (2006), and Ives and Patterson (2008). The auditory image is similar in form to the autocorrelogram (Slaney and Lyon 1990; Meddis and Hewitt 1991; Yost 1996) but the construction of the auditory image is more efficient and it preserves the temporal asymmetry of pulse-resonance sounds. The similarities and differences between auditory images and autocorrelograms are described in Patterson and Irino (1998).

2.4.2 *The Spectral Profile and S_f*

While the processing of pulse-resonance sounds up to the level of our initial perception of them may seem complicated, the relationship between the acoustic properties of these sounds, as observed in their waves and log-frequency spectra, and the features that appear in the auditory images of pulse-resonance sounds is relatively straightforward. In Fig. 2.2, the spectral profile to the right of each auditory image is the average of the activity in the image across time intervals; it simulates the tonotopic distribution of activity observed in the cochlea and in neural centers of the auditory pathway up to auditory cortex. The frequency axis is quasi-logarithmic like the tonotopic dimension of the cochlea (Moore and Glasberg 1983). The spectral profile of the auditory image is similar to the "excitation pattern" described by, for example, Zwicker (1974) and Glasberg and Moore (1990), inasmuch as they all simulate the distribution of activity along the tonotopic axis in the auditory system with a compressed measure of magnitude.

The three peaks in the spectral profile of the /a/ (G2) of the baritone in Fig. 2.2a show the formants of this vowel. Note, that the profile from AIM is similar to the envelope of the magnitude spectrum of the child's vowel, shown in Fig. 2.1b, except that the pattern in Fig. 2.2a is shifted toward the origin with respect to that in Fig. 2.1b because in Fig. 2.2a, the singer is an adult.

The spectral profile of the auditory image is similar in form to the envelope of the magnitude spectrum. Both are *covariant* representations of family and register information (van Dinther and Patterson 2006); the family information is contained in the

shape of the envelope, and the register information is in the position of the envelope, S_f , along the frequency axis. Comparison of the spectral profiles of the auditory images in Fig. 2.2a and b show that, whereas the spectral envelope of the voice is characterized by three distinct peaks, or formants, the envelope of the French horn is characterized by one broad region of activity.

2.4.3 *The Time-Interval Profile and S_s*

The resolution of the auditory filter, at the sound levels where we normally listen to music, is not sufficient to define individual harmonics of pulse-resonance sounds beyond the first three or four harmonics (e.g., Ives and Patterson 2008). As a result, the fine structure of the magnitude spectrum and S_s are not readily apparent in the spectral profile of the auditory image for musical sounds. However, the S_s information is present in the auditory image, in the form of the vertical ridge in the 10-ms region of the image. The ridge shows that there is a concentration of activity at the period of the tone in most channels of the auditory images in Fig. 2.2a and b. Thus, the acoustic scale of the source is readily observed in this simulation of the neural representation of sound, even though the construction of the auditory image includes a temporal integration process with a half life of 30 ms. This is because strobed temporal integration preserves the temporal fine structure of periodic components of sounds like the sustained parts of vowels and musical notes.

Moreover, the temporal information associated with the acoustic scale of the source is enhanced in the time-interval profile of the auditory image. This profile appears below the auditory image and shows the activity averaged across filter channels. In this time-interval profile, the position of the largest peak (in the region to the left of 1.25 ms) provides an accurate estimate of the period of the sound (for G2, 10.2 ms). Moreover, the height of the peak, relative to the level of the background at the foot of the peak, provides a good measure of the salience of the pitch percept (Yost et al. 1996; Patterson et al. 2000; Ives and Patterson 2008). Thus, in time-domain models involving auditory images, the most obvious correlate of the acoustic scale of the source, S_s , in an instrument is a concentration of time intervals at a particular value in the temporal profile. This form of S_s information is more like the time between peaks in the sound wave (Fig. 2.2a) rather than the position of the fine structure in the magnitude spectrum of the sound (Fig. 2.2b).

2.4.4 *Summary of Auditory Image Construction and the Acoustic Scale Information in the Image*

In auditory models of perception, the auditory image that simulates the neural substrate of perception is typically constructed in four stages: A spectral weighting function, similar to the audiogram in form, simulates the middle-ear filtering that

limits sensitivity to very high and very low frequencies. An auditory filterbank simulates the spectral analysis performed in the cochlea. Neural transduction is simulated with half-wave rectification and low-pass filtering. A sophisticated form of temporal integration stabilizes the repeating neural patterns produced by pulse-resonance sounds and completes the construction of the auditory image.

The main vertical ridge in the auditory image, and the corresponding peak in the time-interval profile, are the auditory model's representation of the acoustic scale of the source, S_s . They move left to longer time intervals as the pulse rate of the sound decreases, and to the right to shorter time intervals as the pulse rate increases. When this S_s marker stands out clearly in the time-interval profile well above the background activity, the sound is effectively periodic and the tone is heard to have a strong pitch. When the scale of the filter, S_f , changes, the complex pattern in the auditory image simply moves up or down in frequency *without changing shape*. Similarly, the distribution of activity in the spectral profile of the image moves up or down without changing shape.

2.5 The Acoustic Properties of Pulse-Resonance Sounds and the Auditory Variables of Perception

The final section of this chapter reviews the relationship between the acoustic properties of sound and three variables of auditory perception, loudness, pitch, and timbre, to illustrate how they relate to the variables of music perception described in the sections above, namely, melody, instrument family and register within a family. The American National Standards Institute (ANSI) has provided official definitions of loudness, pitch, and timbre, and these definitions are widely quoted. This section begins with the definitions as they appear in ANSI (1994), as they might have been expected to specify just those relationships between physical and perceptual variables that we require to explain the perception of musical notes. The definitions are:

- 12.03 Loudness. That attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from soft to loud.
- 12.01 Pitch. 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. Note: The pitch of a sound may be described by the frequency or frequency level of that simple tone having a specified sound pressure level that is judged by listeners to produce the same pitch.
- 12.09 Timbre. That attribute of auditory sensation that enables a listener to judge that two nonidentical sounds, similarly presented and having the same loudness and pitch, are dissimilar. Note: Timbre depends primarily upon the frequency spectrum, although it also depends upon the sound pressure and the temporal characteristics of the sound.

These definitions are useful, inasmuch as they illustrate the desire to relate properties of perception to physical properties of sound, and they illustrate what is regarded by auditory scientists as a principled way of proceeding with this task. Unfortunately, the definitions focus on the perceptual properties without, in the end, specifying the relationship of each to the corresponding, acoustic, or physical variables, other than to say that both pitch and timbre *depend primarily on the frequency content of the sound*. Although true, this is not very helpful because it does not say which aspect of the frequency information is associated with pitch and which aspect is associated with timbre. The discussion of acoustic scale in Sect. 2.2 suggests that, for musical sounds at least, we can be more specific about the relationship between the acoustic properties of sound and the perceptions associated with musical notes and instruments. In particular, S_s , the position of the fine structure of the magnitude spectrum, largely determines the pitch of a musical note, and a melody is an ordered sequence of S_s values. The shape of the spectral envelope is closely associated with the perception of instrument family, or the family aspect of timbre. So it is envelope shape that supports the general distinction between, for example, brass and string instruments. And, S_f , the position of the envelope of the magnitude spectrum, combines with S_s to determine the register of the instrument within a family. The acoustic scale variables S_s and S_f are also prime determinants of our perception of the size of an instrument or the height of a singer. In this final section of the chapter, we review the relationship between these acoustic properties of sound and the traditional auditory variables, pitch and timbre, with a view to developing a more useful description of the mapping between the acoustic and auditory variables as they pertain to music perception.

2.5.1 *Effect of Source Size on Pitch and Timbre*

Consider the definitions of pitch and timbre, and the question of how we perceive the physical changes that take place in a vowel as a child grows up, or how we perceive the physical changes that take place in a musical note as it is played on successively larger members of an instrument family, for example, when a trumpet, trombone, and tuba play C3, one after another. The logic of the ANSI definition of timbre is not entirely clear, but it would appear to involve a process of elimination, in which variables of auditory perception that *do not* affect timbre are identified and separated from the remaining variables, which by default are part of timbre. The perceptual variables of particular interest are duration, loudness and pitch.

Duration is the variable that is most obviously separable from timbre, and it illustrates the logic underlying the definition of timbre (although there is not actually a standard definition of the perception of duration). If a singer holds a note for a longer rather than a shorter period, it produces a discriminable change in the sound but it is not a change in timbre. Duration has no effect on the magnitude spectrum of a sound, once the duration is well beyond that of the temporal window used to produce the magnitude spectrum. The sustained notes of music are typically longer

than 200 ms in duration, and the window used to produce the magnitude spectrum is usually less than 100 ms, so duration is unlikely to play a significant role in family timbre or register timbre. In general, then, the perceptual change associated with a change in the duration of a sustained note is separable from changes in the timbre of the note.

Loudness is also largely separable from timbre. If we turn up the volume control when playing a recording, the change will be perceived predominantly as an increase in loudness. The pitch of any given vowel and the timbre of that vowel will be essentially unaffected by the manipulation. The increase in the intensity of the sound produces a change in the magnitude spectrum of the vowel – both the fine structure and the envelope shift vertically upwards – but there is no change in the frequencies of the components of the fine structure and there is no change in the relative amplitudes of the harmonics. Nor is there any change in the shape of the spectral envelope. So, loudness is also separable from timbre.

Thus, acoustic variables that do not affect either the shape of the envelope of the magnitude spectrum or the frequencies of the spectral components do not affect the timbre of the sound. The question is: “What happens when a simple shift is applied to the position of the fine structure, or to the position of the envelope, of a sound (on a log-frequency axis), that is, when we change S_s , S_f , or both?” The current definition of timbre suggests that a change in S_s , which is heard as a change in pitch, does not affect the timbre of the sound, whereas a change in S_f , which is heard as a change in speaker size or instrument size, does affect the timbre of the sound. This is where the current definition of timbre becomes problematic, that is, when it treats the two aspects of acoustic scale differently with regard to their role in the perception of timbre.

Note, in passing, that shifting the position of the fine structure of the magnitude spectrum, while holding the envelope fixed, produces large changes in the relative magnitudes of the harmonics as they move through the region of a formant peak. So the relative magnitudes of the components in the spectrum can change substantially without producing a change in timbre, by the current definition. Note, also, that shifting the envelope of the magnitude spectrum while holding the position of the fine-structure constant produces similar changes in the relative magnitudes of the component frequencies as they move through formant regions. Such shifts do not change the timbre category of a musical sound (the family timbre); they change the apparent size of the source, and if the change is large enough they change the perceived register of the instrument, which, of course, is a timbre change, by the current definition.

2.5.2 Acoustic Scale “Melodies” and the Perception of Pitch and Timbre

The discussion focuses on a set of four melodies designed to emphasize the role of the acoustic scale variables in the perception of vocal pitch and timbre. The novel



Fig. 2.4 Musical notation for four short melodies. The black notes show the acoustic scale of the source, S_s , and thus, the melodic pitch during the course of the musical sequence. The gray, flipped, notes represent the acoustic scale of the filter, S_f , on a musical scale. The original speaker's voice defines the note E (the bottom line on the staves) for both acoustic scales

aspect of the melodies is that, in some cases, the acoustic scale of the filter, S_f , varies over the course of the melody, either on its own, or in conjunction with changes in S_s . The scale of the filter is normally fixed when an instrument plays a melody. A form of musical notation for the melodies is presented in Fig. 2.4; it shows that the melodies all have four bars containing a total of eight notes. The melodies are in $\frac{3}{4}$ time, with the fourth and eighth notes extended to give the sequence a musical feel. The melodies have a “phonological text,” that is, the notes are sung as syllables (*pi, pe, ko, kuuu; ni, ne, mo, muuu*), which emphasizes the human quality of the voice. As the timbre changes from vowel to vowel, it engages the phonological system and allows us to distinguish the role of envelope shape in melody perception, from the role of S_s and the role of S_f . The phonological text is the same for all four melodies.

The syllables were originally sung by an adult male (author R. P.) who has an average GPR of about 120 cps and a vocal tract length of about 16.5 cm. STRAIGHT (Kawahara and Irino 2004) was used to vary the scale of the source, S_s and the scale of the filter, S_f , for each of the syllables, to simulate changes in the GPR and VTL of the singer. The matrix of tones used to produce the melodies is shown in Fig. 2.5. The abscissa of the matrix (x -axis) is the acoustic scale of the source, S_s , and it was varied to produce an octave of notes using the diatonic major scale of Western music. The ordinate of the matrix (y -axis) is the acoustic scale of the filter, S_f , and it was varied to simulate voices with an octave range of vocal tract

lengths ranging from about 10–20 cm. As with the S_s dimension, the specific values of S_f were determined by the diatonic major scale of Western music. In other words, the S_f ratio between any two notes has the same numerical value as the corresponding S_s ratio, and the values of the S_f ratios are indicated in musical notation by the note names, C, D, E, etc. The manipulation of S_f effectively extends the domain of notes from a diatonic musical scale to a diatonic musical plane as shown in Fig. 2.5.

The arrows in Fig. 2.5 show the sequences of notes in each melody. This alternative notation for the melodies illustrates the interaction of the acoustic scale variables. Returning to Fig. 2.4, for each melody, the black notes show the progression of intervals for S_s (or GPR) as each melody proceeds, and the grey notes show the progression of intervals for S_f (or VTL) as the melody proceeds. The sound files for the melodies are available at <http://www.acousticscale.org/link/SHAR2010Demo>. The shaded note (E, E) on the S_s – S_f plane provides the anchor for the notation; it has the same GPR and VTL values as the original syllables.

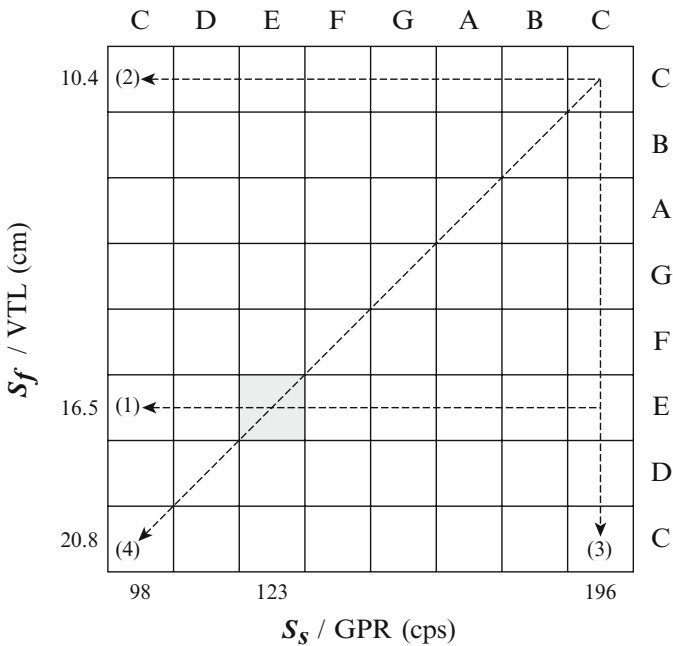


Fig. 2.5 The S_s – S_f plane, or GPR–VTL plane. The abscissa is the acoustic scale of the source S_s , increasing from left to right over an octave. The ordinate is the acoustic scale of the filter S_f doubling from top to bottom. The plane is partitioned into squares that represent the musical intervals. The square associated with the original speaker is highlighted in gray. The dashed lines show the progression of notes in the four melodies of Fig. 2.4

2.5.2.1 Melody 1

The first melody simulates the normal situation wherein a singer with a fixed vocal tract length (VTL) varies the tension of the vocal cords to vary S_s in accordance with the black notes in Staff (1) of Fig. 2.4. The gray notes (for S_f) do not vary in this melody, indicating that the VTL of the singer is fixed. The VTL is relatively long, so the singer is heard to be an adult male. The pitch of the voice drops by an octave over the course of the melody from about 200 cps, which is well above the original pitch, down to about 100 cps, which is a few notes below the original pitch. This descending melody is within the normal range for a tenor, and the melody sounds natural. As the melody proceeds, the fine-structure of the spectrum, S_s , shifts, as a unit, with each change in GPR, and over the course of the melody, it shifts an octave towards the origin. The ANSI definition of timbre implies that these relatively large S_s changes, which produce large pitch changes, do *not* produce timbre changes, and this seems entirely compatible with what we hear in this melody. So, this melody illustrates the commonly held belief, embodied in the ANSI definitions, that pitch is largely separable from timbre, much as duration and loudness are.

2.5.2.2 Melody 2

Problems arise when we extend the example and synthesize a version of the same melody but with a singer that has a much shorter vocal tract, like that of a small child (Fig. 2.4, Staff [2]). There is no problem at the start of the melody; it just sounds like a child singing the melody. The starting pitch is low for the voice of a small child but not impossibly so. As the melody proceeds, however, the pitch decreases by a full octave, which takes it beyond the normal range for a child. As a result, in the latter part of the melody, we hear the voice quality change and, by the end of the melody, the child comes to sound rather more like a dwarf. The ANSI definition of timbre does not provide any basis for understanding the voice quality change from a child to a dwarf; within the traditional framework, the changes that we hear as the melody proceeds are just pitch changes. But traditionally, voice quality changes associated with a change in speaker changed are regarded as timbre changes. This is the first form of problem with the standard definition of timbre – changes that are nominally pitch changes producing what would normally be classified as a timbre change.

2.5.2.3 Melody 3

In the next example (Fig. 2.4, Staff [3]), the roles of the acoustic-scale variables, S_s and S_f , are reversed. The position of the fine structure, S_s , is held fixed while the position of the envelope, S_f , shifts by an octave toward the origin. The change in S_f

simulates a doubling of the VTL, from about 10–20 cm, which would normally be associated with a doubling of height. The S_f ratios between successive notes of the melody have the same numerical values as the S_s ratios of the first two melodies. As melody 3 proceeds and the envelope shifts down by an octave, the child seems to get ever larger, the voice comes to sound something like that of a counter tenor, that is, a tall person with an inordinately high pitch. The ANSI definition of timbre does not say anything specific about how changes in the position of the spectral envelope affect timbre or voice quality; the acoustic scale variable, S_f , was not recognized when these standards were written. Nevertheless, the definition gives the impression that any change in the spectrum that produces an audible change in the perception of the sound, without producing a change in duration, loudness or pitch, produces a change in timbre. Experiments with scaled vowels and syllables show that the just noticeable change in S_f is about 7% for vowels (Smith et al. 2005) and 5% for syllables (Ives et al. 2005), so the intervals in the melody would be expected to produce perceptible S_f changes. Because traditionally, voice quality changes are thought to be timbre changes, the fact that the singer at the start of the melody (a child) is different from the singer at the end of the melody (a counter tenor) seems compatible with the definition of timbre; the singer changes and the timbre changes. However, we are left with the problem that large changes in S_s and S_f both seem to produce changes in voice quality, but whereas the perceptual changes associated with large shifts of the fine-structure along the log-frequency axis are not timbre changes, the perceptual changes associated with large shifts of the envelope along the same log-frequency axis are timbre changes, according to the ANSI definition. They both produce changes in the relative amplitudes of the spectral components, but neither changes the shape of the envelope and neither form of shift alters the phonological values of the individual syllables.

2.5.2.4 Melody 4

The problems involved in attempting to unify the perception of voice quality with the definition of timbre become more complex when we consider melodies where both S_s and S_f change as the melody proceeds. Consider the melody produced by covarying S_s and S_f to produce the notes along the diagonal of the S_s – S_f plane (Fig. 2.5). The musical notation for the melody is shown in Fig. 2.4, Staff (4). This melody is perceived to descend an octave as the sequence proceeds, and there is a progressive increase in the perceived size of the singer from a child to an adult male (with one momentary reversal at the start of the second phrase). It is as if we had a set of singers varying in age from 4–18 in a row on stage, and we had them each sing their assigned syllable in order, and in time, to produce the melody. This melody, in combination with the others, makes it clear that there is an entire plane of singers with different vocal qualities defined by different combinations of the acoustic scale variables, S_s and S_f . The realization that there is a whole plane of voice qualities makes it clear just how difficult it would be to produce a clean

definition of timbre that excludes one of the acoustic scale variables, S_s , and not the other, S_f . If changes in voice quality are changes in timbre, then changes in pitch (S_s) can produce changes in timbre. This would seem to undermine the utility of the current definitions of pitch and timbre.

2.5.3 *The Concept of Acoustic Scale and the Definitions of Pitch and Timbre*

2.5.3.1 The “Second Dimension of Pitch” Hypothesis

At first glance, there would appear to be a fairly simple way to solve the problem; we could designate the perceptual dimension associated with the acoustic scale of the filter, S_f , to be a second dimension of pitch. Then, this second dimension of pitch could be excluded from the definition of timbre along with the first dimension of pitch. For the singing voice, manipulation of S_f on its own would sound like the change in perception that occurs over the course of melody 3, where S_s is fixed on the upper C and S_f decreases by a factor of two over the course of the melody. This does, however, lead to several problems. First, semitone changes in the scale of the filter, S_f , are barely large enough to hear differences in the associated perception so this second dimension of pitch would not support accurate perception of novel melodies, in the way that the first dimension of pitch does (e.g., Pressnitzer et al. 2001; Ives and Patterson 2008). The salience of changes in S_f is more like the salience of the weak S_s pitch that arises when the energy in a tone is restricted to high, unresolved harmonics, and pitch discrimination requires S_s changes of four semitones, or more. The second dimension of pitch would, in some sense, satisfy the ANSI definition of pitch which is not concerned with melodies, and which only requires that the attribute of auditory sensation can be used to order notes on a scale extending from low to high. It seems reasonable to say that the tones at the start of melody 3 sound “higher” than the tones at the end of the melody, which would support the “second dimension of pitch” hypothesis.

The “second dimension of pitch” hypothesis also leads to another problem. To determine the pitch of a sound, it is traditional to match the pitch of that sound to the pitch of either a sinusoid or a click train, that is, to a perception that is based on the scale of the source, S_s . Moreover, it seems likely that if listeners were asked to pitch match each of the notes in melody 3, among a larger set of sounds that diverted attention from the orderly progression of S_f in the melody, they would probably match all of the tones with the same sinusoid or the same click train, and the pitch of the matching stimulus (bound to an S_s value) would be the upper C. This would leave us with the problem that the second dimension of pitch, based on S_f , changes the perception of the sound but it does not change the pitch to which the sound is matched (its S_s value). So the “pitch” change associated with a change in S_f would have to be segregated from a normal pitch change and given a separate definition. It would also require changes in the ANSI definitions of pitch and timbre because currently, a change in perception (such as that associated with changes in S_f) that

does not produce a change in S_s pitch (or loudness, or duration) is a change in timbre. In short, the “second dimension of pitch” hypothesis would appear to lead us back to the position that changes in S_f produce changes in the timbre of the sound.

The “second dimension of pitch” hypothesis also implies that if we play a random sequence of notes on the musical plane of Fig. 2.5, the voice quality changes that we hear are all pitch changes, and they involve no change in timbre. This seems unreasonable when the acoustic scale changes are sufficiently large to produce a clear change in the perception of *who* is singing.

Finally, there is the problem that many people hear the perceptual change in melody 3 as a change in speaker size, and they hear a more pronounced change in speaker size when changes in S_f are combined with changes in S_s , as in melody 4. To ignore the perception of speaker size, is another problem inherent in the “second dimension of pitch” hypothesis; source size is an important aspect of perception, and pretending that changes in the perception of source size are just pitch changes seems like a fundamental mistake for a model of perception.

2.5.3.2 The Scale of the Filter, S_f , as a Dimension of Timbre

Rather than co-opting the acoustic scale of the filter, S_f , to be a second dimension of pitch, it would seem more reasonable to think of it as an internal dimension of timbre – a dimension of timbre that for voices is associated with vocal register, singer gender, and singer size. This, however, leads to a different problem which is, in some sense, the inverse of the “second dimension of pitch” problem. Once it is recognized that shifting the position of the fine structure of the spectrum is inherently similar to shifting the position of the envelope of the spectrum, and that the two position variables are different aspects of the same property of sound (acoustic scale), then it seems unreasonable to have one of these variables, S_f , within the realm of timbre and the other, S_s , outside the realm of timbre. For example, consider the issue of voice quality; both of the acoustic scale dimensions affect voice quality and they interact in the production of a specific voice quality (e.g., man, woman, child, dwarf, counter tenor). Moreover, the scale of the source, S_s , affects the perception of the singer’s size, in a way that is similar to the perceptual effect of the scale of the filter, S_f (Smith and Patterson 2005). Thus, if we define the scale of the filter, S_f , to be a dimension of timbre, then we need to consider that the scale of the source, S_s , may also need to be a dimension of timbre. After all, large changes in S_s affect voice quality which is normally considered to be an aspect of timbre.

2.5.4 The Independence of Spectral Envelope Shape

There is one further aspect of the perception of these melodies that should be emphasized, which is that neither of the acoustic scale manipulations causes a

change in the perception of the phonology of the syllables; we always hear *pi*, *pe*, *ko*, *kuuu*; *ni*, *ne*, *mo*, *muuu*, independent of the VTL and GPR values of the singers. That is, the changes in timbre that give rise to the perception of a sequence of syllables are unaffected by changes in S_s and S_f , even when these scale changes are large (Ives et al. 2005; Smith et al. 2005). The changes in timbre that define the phonology are associated with changes in the shape of the envelope, as opposed to the position of the spectral envelope or the position of the spectral fine structure. Changes in the shape of the envelope produce changes in vowel type in speech and changes in instrument family in music. Changing the position of the envelope and changing the position of the fine structure both produce substantial changes in the relative amplitudes of the components of the magnitude spectrum, but they do not change the timbre category of these sounds; that is, they do not change the vowel type in speech or the instrument family in music.

2.5.5 Summary

The ANSI definitions of pitch and timbre are not much help in understanding the perception of musical tones, in the sense of understanding what gives rise to the perception of melody, instrument family, and register within a family. The ANSI definitions simply associate both pitch and timbre with unspecified aspects of the frequency content of a sound. In music and speech research, it is traditional to segregate one aspect of the frequency information, namely, F0 (the repetition rate of the sound), from the remainder of the information which is represented by the spectrogram. F0 is then associated with the pitch of the instrument or the pitch of the voice, in the same way that we have associated the scale of the source, S_s , with pitch. Thus, in music and speech research there is, at least, the segregation of the main determinant of pitch from the distribution of frequency information across the acoustic frequency dimension. The difference between these approaches and the acoustic-scale approach presented in this chapter is illustrated in Fig. 2.6. The upper row shows how the frequency information is (or is not) divided up in each case, and

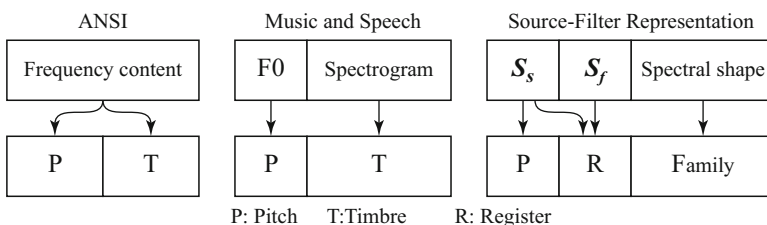


Fig. 2.6 The relationship between the acoustic variables (upper row) and the psychological variables related to the perception of a musical tone (lower row) as defined in the ANSI standard (ANSI 1994), in music and speech research, and using the acoustic scale variables

the lower row shows the components of auditory perception; the arrows indicate the associations between the components of the frequency information and the components of perception. In the first column, which corresponds to the ANSI definition, there is only one arrow associating all of the frequency content, indiscriminately, with both pitch and timbre. The second column, corresponding to music and speech research, shows how F0 is segregated from the spectrogram and associated with pitch.

The third column shows how the scale of the source, S_s , and the scale of the filter, S_f , are segregated from the shape of the envelope of the magnitude spectrum in the current approach. The scale of the source is directly related to musical pitch and melody. The shape of the envelope is directly related to the family aspect of timbre, and for the human voice, this is further subdivided into different vowel types. These aspects of the mapping from acoustic properties to perceptual variables are straightforward. The mapping between acoustic properties and register within a family is a little more complicated; both of the acoustic scale variables contribute to the perception of *register*. Both of the acoustic scale variables also contribute to the perception of instrument size and singer size, which are related perceptions in different contexts. It is also the case that the relative magnitude of the acoustic scale variables contributes to our perception of whether a specific instrument is a good, or bad, example of its class. Although the division of frequency information into three components, and the mapping from these components to the perception of musical tones, is somewhat more complicated than in traditional descriptions, it is not excessively complicated, and it does provide for a much better understanding of how the physical properties of instruments, and the acoustic properties of sound relate, to the auditory perceptions that musical tones produce.

2.6 Conclusions

Recent research on the role of acoustic scale in the perception of sound suggests that the frequency information observed in the magnitude spectrum of a sound is segregated by the auditory system into three parts: the spectral envelope shape, the acoustic scale of the source, S_s , and the acoustic scale of the filter, S_f . The spectral envelope shape determines the basic timbre category of a sound, which in music is the instrument family, and in the singing voice expands to produce the different vowel types. These timbre categories are largely independent of the acoustic scale variables, S_s and S_f . In speech, these two acoustic scale variables jointly determine much of the static voice quality of the speaker, and thus our perception of a speaker's sex and size (e.g., Smith and Patterson 2005). This suggests that it would be useful to distinguish between the "what" and "who" of timbre in speech, that is, what is being said, and who is saying it. With regard to the timbre of musical tones, the distinction between envelope shape and the acoustic scale variables provides an

explanation for the distinction between family timbre (envelope shape) and register timbre (S_s and S_p). In both speech and music, S_s exhibits a limited degree of independence from timbre inasmuch as (1) variation of GPR to produce prosodic distinctions does not change the perception of who is speaking, and (2) variation of the pulse rate in musical instruments to produce a melody does not change the perception of the instrument that is playing. There are, however, limits to the independence; large changes in pulse rate produce changes in the perception of who is speaking or which member of an instrument family is playing.

Acknowledgments The authors were supported by the UK Medical Research Council (G0500221; G9900369) during the preparation of this chapter. They thank Jim Woodhouse for useful discussions on the production of notes by the violin, and on acoustic scaling in the string family.

References

- ANSI (1994) American National Standard Acoustical Terminology, ANSI S1.1–1994 (R1999). New York: American National Standard Institute.
- Benade AH (1976) *Fundamentals of Musical Acoustics*. Oxford: Oxford University Press.
- Benade AH, Lutgen SJ (1988) The saxophone spectrum. *J Acoust Soc Am* 83:1900–1907.
- Chiba T, Kajiyama M (1941) *The Vowel, its Nature and Structure*. Tokyo: Tokyo-Kaiseikan.
- Cohen L (1993) The scale representation. *IEEE Trans Sig Proc* 41:3275–3292.
- de Cheveigné A (2005) Pitch perception models. In: Plack CJ, Oxenham AJ, Fay RR, Popper AN (eds), *Pitch: Neural Coding and Perception*. New York: Springer, pp. 169–233.
- Fant G (1960) *Acoustic Theory Of Speech Production*. The Hague: Mouton De Gruyter.
- Fitch WT, Giedd J (1999) Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *J Acoust Soc Am* 106:1511–1522.
- Fitch WT, Reby D (2001) The descended larynx is not uniquely human. *Proc R Soc Lond Ser B* 268:1669–1675.
- Fletcher NH (1978) Mode locking in nonlinearly excited inharmonic musical oscillators. *J Acoust Soc Am* 64:1566–1569.
- Fletcher NH, Rossing TD (1998) *The Physics of Musical Instruments*. New York: Springer.
- Glasberg BR, Moore BCJ (1990) Derivation of auditory filter shapes from notched-noise data. *Hear Res* 47:103–138.
- Glasberg BR, Moore BCJ (2002) A model of loudness applicable to time-varying sounds. *J Audio Eng Soc* 50:331–342.
- Helmholtz HLF (1875) *On the Sensations of Tone as a Physiological Basis for the Theory of Music*. London: Longmans, Green.
- Hutchins CM (1967) Founding a family of fiddles. *Phys Today* 20:23–37.
- Hutchins CM (1980) The new violin family. In: Benade AH (ed), *Sound Generation in Winds, Strings, Computers*. Stockholm: The Royal Swedish Academy of Music, pp. 182–203.
- Irino T, Patterson RD (2001) A compressive gammachirp auditory filter for both physiological and psychophysical data. *J Acoust Soc Am* 109:2008–2022.
- Irino T, Patterson RD (2006) A dynamic compressive gammachirp auditory filterbank. *IEEE Trans Audio Speech Lang Processing* 14:2222–2232.
- Ives DT, Patterson RD (2008) Pitch strength decreases as F0 and harmonic resolution increase in complex tones composed exclusively of high harmonics. *J Acoust Soc Am* 123:2670–2679.
- Ives DT, Smith DRR, Patterson RD (2005) Discrimination of speaker size from syllable phrases. *J Acoust Soc Am* 118:3186–3822.

- Kawahara H, Irino T (2004) Underlying principles of a high-quality speech manipulation system STRAIGHT and its application to speech segregation. In: Divenyi PL (ed), *Speech Separation by Humans and Machines*. Kluwer Academic, pp. 167–180.
- Kennedy M (1985) *The Oxford Dictionary of Music*. Oxford: Oxford University Press.
- 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 RD, Nobbe A, Fastl H (2003) Microsecond temporal resolution in monaural hearing without spectral cues? *J Acoust Soc Am* 113:2790–2800.
- Lee S, Potamianos A, Narayanan S (1999) Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *J Acoust Soc Am* 105:1455–1468.
- Licklider JCR (1951) A duplex theory of pitch perception. *Experientia* 7:128–133.
- McIntyre ME, Schumacher RT, Woodhouse J (1983) On the oscillations of musical instruments. *J Acoust Soc Am* 74:1325–1345.
- 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.
- Molin NE, Lindgren L-E, Jansson EV (1988) Parameters of violin plates and their influence on the plate modes. *J Acoust Soc Am* 83:281–291.
- Moore BCJ, Glasberg BR (1983) Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *J Acoust Soc Am* 74:750–753.
- 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.
- Patterson RD, Irino T (1998) Modeling temporal asymmetry in the auditory system. *J Acoust Soc Am* 104:2967–2979.
- Patterson RD, Robinson K, Holdsworth J, McKeown D, Zhang C, Allerhand M (1992) Complex sounds and auditory images. In: Cazals Y, Demany L, Horner K (eds), *Auditory Physiology and Perception*. Oxford: Pergamon Press, pp. 429–446.
- 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.
- Patterson RD, Yost WA, Handel S, Datta AJ (2000) The perceptual tone/noise ratio of merged iterated rippled noises. *J Acoust Soc Am* 107:1578–1588.
- Patterson RD, Unoki M, Irino T (2003) Extending the domain of center frequencies for the compressive gammachirp auditory filter. *J Acoust Soc Am* 114:1529–1542.
- Patterson RD, Smith DDR, van Dinther R, Walters TC (2008) Size Information in the production and perception of communication sounds. In: Yost WA, Popper AN, Fay RR (eds), *Auditory Perception of Sound Sources*. New York: Springer, pp. 43–75.
- Peterson GE, Barney HL (1952) Control methods used in a study of the vowels. *J Acoust Soc Am* 24:175–184.
- Pressnitzer D, Patterson RD, Krumbholtz K (2001) The lower limit of melodic pitch. *J Acoust Soc Am* 109:2074–2084.
- Schelleng JC (1963) The violin as a circuit. *J Acoust Soc Am* 35:326–338.
- Slaney M, Lyon RF (1990) A perceptual pitch detector. In: *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Albuquerque, New Mexico, pp. 357–360.
- Smith DRR, Patterson RD (2005) The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *J Acoust Soc Am* 118:3177–3186.
- Smith DRR, Patterson RD, Turner RE, Kawahara H, Irino T (2005) The processing and perception of size information in speech sounds. *J Acoust Soc Am* 117:305–318.
- Sprague MW (2000) The single sonic muscle twitch model for the sound-production mechanism in the weakfish, *Cynoscion regalis*. *J Acoust Soc Am* 108:2430–2437.
- Turner RE, Walters TC, Monaghan JJM, Patterson RD (2009) A statistical formant-pattern model for estimating vocal-tract length from formant frequency data. *J Acoust Soc Am* 125:2374–2386.

- Unoki M, Irino T, Glasberg B, Moore BC, Patterson RD (2006) Comparison of the roex and gammachirp filters as representations of the auditory filter. *J Acoust Soc Am* 120:1474–1492.
- van Dinther R, Patterson RD (2006) Perception of acoustic scale and size in musical instrument sounds. *J Acoust Soc Am* 120:2158–2176.
- Yost WA (1996) Pitch of iterated rippled noise. *J Acoust Soc Am* 100:511–518.
- Yost WA (2009) Pitch perception. *Atten Percept Psychophys* 71:1701–1715.
- 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 RD, Sheft S (1998) The role of the envelope in processing iterated rippled noise. *J Acoust Soc Am* 104:2349–2361.
- Zwicker E (1974) On the psychophysical equivalent of tuning curves. In: Zwicker E, Terhardt E (eds), *Facts and Models in Hearing*. New-York: Springer-Verlag, pp. 132–140.

Chapter 3

A Theory of Tonal Hierarchies in Music

Carol L. Krumhansl and Lola L. Cuddy

3.1 Introduction

One of the most pervasive structural principles found in music historically and cross-culturally is a hierarchy of tones. Certain tones serve as reference pitches; they are stable, repeated frequently, are emphasized rhythmically, and appear at structurally important positions in musical phrases. The details of the hierarchies differ across styles and cultures. Variation occurs in the particular intervals formed by pitches in the musical scale and the hierarchical levels assigned to pitches within the scale. This variability suggests that an explanation for how these hierarchies are formed cannot be derived from invariant acoustic facts, such as the harmonic structure (overtones) of complex tones. Rather, the evidence increasingly suggests that these hierarchies are products of cognition and, moreover, that they rely on fundamental psychological principles shared by other domains of perception and cognition.

In this chapter, a theory of tonal hierarchies is presented that rests upon three interrelated propositions. The first is that tonal hierarchies have psychological reality. The first is that tonal hierarchies have psychological reality – that is, they are represented cognitively and play that is, they are represented cognitively and play a central role in how musical sequences are perceived, organized, and remembered and in how expectations are formed during listening. This proposition implies that effects of tonal hierarchies should surface in a variety of empirical measures, such as direct judgments of musical structure, memory errors, and neurophysiological measures. The second proposition is that the tonal hierarchies are also musical facts. As such, it is expected that these hierarchies will manifest in the way music is written and how its structure is codified in music theory. Tonal hierarchies should be evident in the musical surface and characterize otherwise diverse musical styles.

C.L. Krumhansl (✉)

Department of Psychology, Cornell University, Ithaca, NY 14853, USA

e-mail: clk4@cornell.edu

L.L. Cuddy (✉)

Department of Psychology, Queen's University, Kingston, Ontario K7L 3N6, Canada

e-mail: Lola.Cuddy@queensu.ca

The third proposition is that statistically frequent patterns in the music should, in most cases, be reliable guides to the listener for abstracting the tonal hierarchy. This proposition would predict that listeners are able to orient relatively rapidly to the style-appropriate tonal hierarchy, and that perceptual judgments should converge with statistical distributions of tones and tone combinations. The psychological challenge with which the present chapter is concerned is the isolation, direct measurement, and quantification of tonal hierarchies.

3.2 Tonal Hierarchy

Tonal hierarchy refers to both a fundamental theoretical concept in describing musical structure and a well-studied empirical phenomenon. As a theoretical concept, the essential idea is that a musical context establishes a hierarchy of tones. Certain musical tones are more prominent, stable, and structurally significant than others, thus yielding a hierarchical ordering of tones. For Western tonal-harmonic music (the prevalent music of the eighteenth and nineteenth centuries) the first tone of the scale (the *tonic*) is said to head the hierarchy. This tone is followed by the fifth and third scale degrees (the *dominant* and the *mediant*, respectively), the other (major or minor) scale tones, and finally the nonscale tones. This hierarchy reflects the influence of triadic (chord) structure in this style, in which consonant chords predominate (for summaries of these elementary aspects of Western music theory, see Handel 1989; Krumhansl 1990a; Patel 2008; Thompson 2008).

As an empirical phenomenon, tonal hierarchy has been extensively investigated in psychological experiments. These investigations are motivated by two general objectives. The first is to test the psychological reality of music-theoretical descriptions. Do the observations made by music theorists about tonal hierarchies have consequences for how musical pitch is perceived and remembered? The second objective is to locate the observed empirical phenomenon within the broader theoretical and methodological framework of psychology. Through what processes are tonal hierarchies internalized? What relationship do tonal hierarchies have to other psychological phenomena, and what techniques of analysis and modeling might clarify the empirical results?

Much of the empirical work reviewed in this chapter has been conducted within the context of the Western tonal-harmonic style. Across musical styles and cultures, however, the notion of *tone (or pitch) centrality* can also be found – that is, one central tone anchors a subset of hierarchically related tones. Moreover, it is possible for an individual piece of music to establish its own unique hierarchy. Findings that point to listeners' sensitivity to such hierarchies are described. Our conceptualization of the tonal hierarchy therefore invokes a pan-stylistic approach to knowledge acquisition and representation of musical structure.

Psychological research on tonal hierarchies developed beginning in the late 1970s as part of an increasing appreciation of the role of cognition in music. The alternative approach up to that time, with a tradition dating to the ancient Greek

philosophers, focused instead on music acoustics. The basic idea was that the formation of musical structures such as scales and chords could be accounted for by the harmonic structure of complex periodic sounds. The cognitive approach, in contrast, sought to understand the role of experience within the musical culture. It raised a host of interrelated questions, including the psychological processes and neural mechanisms involved in learning musical patterns, the role of development and training, and cross-cultural comparisons. The cognitive approach also encouraged the development of quantitative models of music learning and perception.

3.2.1 *Psychological Principles Underlying Tonal Hierarchies*

The structure of tonal hierarchies appears to rely on two basic cognitive principles. The first is the existence of cognitive reference points (Rosch and Mervis 1975; Rosch 1975, 1978, 1979), which motivated the initial empirical studies of tonal hierarchies (Krumhansl 1979; Krumhansl and Shepard 1979; Krumhansl and Kessler 1982). Within categories, certain perceptual and conceptual objects, called *cognitive reference points*, have special psychological status. They are reference points in relation to which other category members are encoded, described, and remembered. In Rosch's work they are sometimes referred to as *prototypes*, although this term seems less apt when applied to musical pitch. Their existence serves the purpose of *cognitive economy* – that is, an internal coding best suited for making distinctions relevant to the domain in question at the same time conserving finite cognitive resources. Empirical work has been performed on cognitive reference points or prototypes in a wide variety of domains, including visual objects, colors, numbers, faces, and personality descriptions. These investigations have shown that cognitive reference points are given priority in processing, are most stable in memory, and have a special role in linguistic descriptions.

We suggest that not only do cognitive reference points function similarly in music, but also they may be especially important there. This is because music does not provide fixed reference tones except as determined by the music itself. Thus, unlike other domains in which cognitive reference points are defined independently of the category (red is perceptually red whether it is or is not thought of in terms of the category of colors), the function of a tone depends entirely on the musical context. Another way to express this is that for most listeners relational processing (relative pitch) predominates over absolute pitch (with pitches having fixed labels independent of context). At a general level, the importance of musical reference points is not merely that they exist, but also that they guide musical perception, memory, thought, and understanding.

The second basic cognitive principle is sensitivity to statistical regularities in music. Statistical regularities that have been considered include the distribution of tones (their frequency of occurrence and their total temporal duration), and the frequency of sequences of tones. Recent research has suggested that statistical learning may play a role in language acquisition (Saffran et al. 1996a, b, 1997).

In this research, infants appear to have learned which syllables frequently co-occur in sequences. A learning process such as this may lead to the identification of combinations of syllables as words. Subsequently, the paradigm has been extended to tones (Saffran et al. 1999; Saffran and Griepentrog 2001). Thus, early in development humans appear to be sensitive to frequent successions of sounds, and this sensitivity may encompass both language and music.

In sum, we propose that regularities within the musical style establish tone centrality. Regularities include repetition of tones and tone sequences, melodic and rhythmic emphasis, durational and metric stress, and positioning of central tones at or near beginnings and endings of phrases. Through repeated exposure to music, listeners implicitly develop a mental representation that captures the regularities. This representation can then be used to encode and remember musical patterns in the future, and generate expectations while listening. Sensitivity to these regularities may also enable listeners to adapt relatively easily to novel musical styles.

3.2.2 *Definitions and Distinctions*

The concept of the tonal hierarchy draws on a long tradition in music theory and history (DeVoto 1986). Various units of musical structure have been abstracted from compositional practice since the seventeenth century and codified. These include scales, modes, chords, keys, and relations among keys (the circle of fifths), described in basic music texts (e.g., Piston 1987). The notion of a tonal hierarchy incorporates relations among all these units in a stable, abstract frame of reference. In this frame of reference all tones and chords are described with respect to the tone that gives the key its name. For example, in C major the first scale tone is C and it is called the *tonic*, and the three-tone chord built on it with the tones C-E-G is called the *tonic triad*. The tone G, which forms a very consonant interval (a fifth) with the tonic, is called the *dominant*, as is the triad G-B-D that is built on it. Similarly, each other tone and chord is designated relative to the tonic.

The tonal hierarchy does not contain information about pitch height. Octave equivalence is assumed. In other words, all members of a pitch class (e.g., C₁, C₂, C₃, C₄ and so on – where the number refers to the octave containing the tone) are represented by a single element (in this case, C). Thus, the hierarchy refers to *pitch classes* rather than to specific pitches. Moreover, the tonal hierarchy does not directly contain information about individual tones as they occur in a musical piece. The order, metric position, timbre, and loudness of tones are not represented. The relative stability of tones in the tonal hierarchy might therefore be characterized as static, independent of the place of the particular tone in the music.

Bharucha (1984) has drawn an important distinction between the tonal hierarchy and the hierarchy created within the framework of a particular piece, or section of music. He named the latter *event hierarchy*; it describes the relative prominence of events in that particular sequence. “Event hierarchies describe the encoding of specific pieces of music; tonal hierarchies embody our tacit or implicit knowledge of the

abstract musical structure of a culture or genre” (Bharucha 1984, p. 421). So, unlike tonal hierarchies that refer to cognitive representations of the structure of music across different pieces of music in the style, event hierarchies refer to a particular piece of music and the place of each event in that piece.

The two hierarchies occupy complementary roles. In listening to music or music-like experimental materials (melodies and harmonic progressions), the listener responds both to the structure provided by the tonal hierarchy and the structure provided by the event hierarchy. Musical activity involves dynamic patterns of stability and instability to which both the tonal and event hierarchies contribute. Understanding the relations between them and their interaction in processing musical structure is a central issue, not yet extensively studied empirically.

3.3 Empirical Research: The Basic Studies

This section outlines the classic findings that illustrate tonal relationships and the methodologies used to establish these findings.

3.3.1 *The Probe Tone Method*

Quantification is the first step in empirical studies because it makes possible the kinds of analytic techniques needed to understand complex human behaviors. An experimental method that has been used to quantify the tonal hierarchy is called the probe-tone method (Krumhansl and Shepard 1979). It was based on the observation that if you hear the incomplete ascending C major scale, C-D-E-F-G-A-B, you strongly expect that the next tone will be the high C. It is the next logical tone in the series, proximal to the last tone of the context, B, and it is the tonic of the key. When, in the experiment, incomplete ascending and descending scale contexts were followed by the tone C (the probe tone), listeners rated it highly as to how well it completed the scale (1 = very badly, 7 = very well). Other probe tones, however, also received fairly high ratings, and they were not necessarily those that are close in pitch to the last tone of the context. For example, the more musically trained listeners also gave high ratings to the dominant, G, and the mediant, E, which together with the C form the tonic triad. The tones of the scale received higher ratings than the nonscale tones, C# D# F# G# and A#. Less musically trained listeners were more influenced by how close the probe tone was to the tone sounded most recently at the end of the context, although their ratings also contained some of the tonal hierarchy pattern.

A subsequent study used this method with a variety of contexts at the beginning of the trials (Krumhansl and Kessler 1982). Contexts were chosen because they are clear indicators of a key. They included the scale, the tonic triad chord, and chord

sequences strongly defining major and minor keys. These contexts were followed by all possible probe tones in the 12-tone chromatic scale, which musically trained listeners were instructed to judge in terms of how well they fit with the preceding context in a musical sense. The results for contexts of the same mode (major or minor) were similar when transposed to a common tonic. Also, the results were largely independent of which particular type of context was used (e.g., chord versus chord cadence). Consequently, the rating data were transposed to a common tonic and averaged over the context types. The resulting values are termed *standardized key profiles*. The values for the major key profile are 6.35, 2.23, 3.48, 2.33, 4.38, 4.09, 2.52, 5.19, 2.39, 3.66, 2.29, 2.88, where the first number corresponds to the mean rating for the tonic of the key, the second to the next of the 12 tones in the chromatic scale, and so on. The values for the minor key context are 6.33, 2.68, 3.52, 5.38, 2.60, 3.53, 2.54, 4.75, 3.98, 2.69, 3.34, 3.17. These are plotted in Fig. 3.1, in which C is assumed to be the tonic. Both major and minor contexts produce clear and musically interpretable hierarchies in the sense that tones are ordered or ranked according to music-theoretic descriptions.

The results of these initial studies suggested that it is possible to obtain quantitative judgments of the degree to which different tones are perceived as stable reference tones in musical contexts. The task appeared to be accessible to listeners who differed considerably in their music training. This was important for further investigations of the responses of listeners without knowledge of specialized vocabularies for describing music, or who were unfamiliar with the musical style. Finally, the results in these and many subsequent studies were quite consistent over a variety of task instructions and musical contexts used to induce a sense of key. Quantification

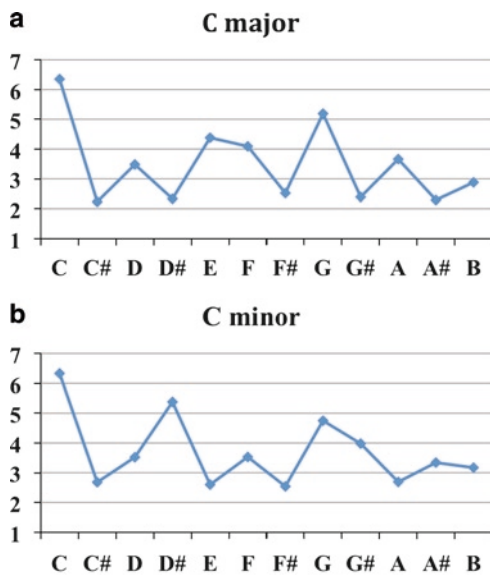


Fig. 3.1 (a) Probe tone ratings for a C major context. (b) Probe tone ratings for a C minor context. Values from Krumhansl and Kessler (1982)

of the tonal hierarchies is an important first step in empirical research but, as seen later, a great deal of research has studied it from a variety of different perspectives.

3.3.2 *Converging Evidence*

To substantiate any theoretical construct, such as the tonal hierarchy, it is important to have evidence from experiments using different methods. This strategy is known as “converging operations” (Garner et al. 1956). This section describes a number of other experimental measures that show influences of the tonal hierarchy. It has an effect on the degree to which tones are perceived as similar to one another (Krumhansl 1979), such that tones high in the hierarchy are perceived as relatively similar to one another. For example, in the key of C major, C and G are perceived as highly related, whereas C# and G# are perceived as distantly related, even though they are just as far apart objectively (in semitones).

In addition, a pair of tones is heard as more related when the second is more stable in the tonal hierarchy than the first (compared to the reverse order). For example, the tones F#-G are perceived as more related to one another than are G-F# because G is higher in the tonal hierarchy than F#. Similar temporal-order asymmetries also appear in memory studies. For example, F# is more often confused with G than G is confused with F# (Krumhansl 1979). These data reflect the proposition that each tone is drawn toward, or expected to resolve to, a tone of greater stability in the tonal hierarchy.

Janata and Reisberg (1988) showed that the tonal hierarchy also influenced reaction time measures in tasks requiring a categorical judgment about a tone’s key membership. For both scale and chord contexts, faster reaction times (in-key/out-of-key) were obtained for tones higher in the hierarchy. In addition, a recency effect was found for the scale context as for the nonmusicians in the original probe tone study (Krumhansl and Shepard 1979).

Miyazaki (1989) found that listeners with absolute pitch named tones highest in tonal hierarchy of C major faster and more accurately than other tones. This is remarkable because it suggests that musical training has a very specific effect on the acquisition of absolute pitch. Most of the early piano repertoire is written in the key of C major and closely related keys. All of these listeners began piano lessons as young as 3–5 years of age, and were believed to have acquired absolute pitch through exposure to piano tones.

The tonal hierarchy also appears in judgments of what tone constitutes a good phrase ending (Palmer and Krumhansl 1987a, b; Boltz 1989a, b). A number of studies show that the tonal hierarchy is one of the factors that influences expectations for melodic continuations (Schmuckler 1989; Krumhansl 1991, 1995b; Cuddy and Lunney 1995; Krumhansl et al. 1999, 2000). Other factors include pitch proximity, interval size, and melodic direction.

The influence of the tonal hierarchy has also been demonstrated in a study of expressive piano performance (Thompson and Cuddy 1997). Expression refers to

the changes in duration and dynamics (loudness) that performers add beyond the notated music. For the harmonized sequences used in their study, the performance was influenced by the tonal hierarchy. Tones that were tonally stable within a key (higher in the tonal hierarchy) tended to be played for longer duration in the melody than those less stable (lower in the tonal hierarchy).

A method used more recently (Aarden 2003, described in Huron 2006) is a reaction-time task in which listeners had to judge whether unfamiliar melodies went up, down, or stayed the same (a tone was repeated). The underlying idea is that reaction times should be faster when the tone conforms to listeners' expectations. His results confirmed this hypothesis, namely, that reaction times were faster for tones higher in the hierarchy. As described later, his data conformed to a very large statistical analysis he did of melodies in major and minor keys.

Finally, tonal expectations result in event-related potentials (ERPs), changes in electrical potentials measured on the surface of the head (Besson and Faïta 1995; Besson et al. 1998). A larger P300 component, a positive change approximately 300 ms after the final tone, was found when a melody ended with a tone out of the scale of its key than a tone in the scale. This finding was especially true for musicians and familiar melodies, suggesting that learning plays some role in producing the effect; however, the effect was also present in nonmusicians, only to a lesser degree.

This section has cited only a small proportion of the studies that have been conducted on tonal hierarchies. A closely related issue that has also been studied extensively is the existence of, and the effects of, a hierarchy of chords. The choice of the experiments reviewed here was to illustrate the variety of approaches that have been taken. Across the studies, consistent effects were found with many different kinds of experimental materials and methods. Thus, the requirement of converging evidence has been satisfied.

3.3.3 Summarizing the Basic Results: Three Principles of Tonal Hierarchies

This consistency across studies enabled the following theoretical summary to be formulated. Bharucha and Krumhansl (1983; see also Krumhansl, 1990a, pp. 140–152) formalized three principles of tonal stability, the relative position of tones in the tonal hierarchy, as a way of summarizing many of the results just described. They are stated in terms of *psychological distance*. If two tones are judged as similar to one another then they are said to be separated by a small psychological distance. Or, another measure of similarity is how often they are confused in memory; if there are many instances of confusion between them then they would be said to have a small psychological distance.

The first principle, *contextual identity*, assumes that not all tones have zero distance from themselves. For example, in a memory task some tones are more often confused

with other tones, whereas others are more often correctly identified as themselves. The principle states that the psychological distance between a tone and itself is smaller (more often remembered and less often confused with other tones) when it is higher in the hierarchy than when it is lower. In the context of C major, for instance, the tone G will be better remembered than the tone F#.

The second principle, *contextual distance*, states that the average perceived distance between two different tones decreases as their position in the hierarchy increases. For example, all else equal, in a C major context, the tones E and G will be judged as closer than the tones F# and A (because E and G are higher in the hierarchy than F# and A) even though their objective distance (in semitones) is the same.

The third principle, *contextual asymmetry*, holds that there will be an effect of the order of two tones. When a tone lower in the hierarchy is followed by one higher in the hierarchy they are perceived as psychologically less distant than when the two tones are played in the opposite order. For example, F# will be perceived as closer to G than G is to F#; the same temporal-order asymmetry would be found in instances of memory confusion. Even more specifically, the size of the order difference will depend on difference in the tones' positions in the tonal hierarchy. For example, the asymmetry between F#-G and G-F# will be larger than the asymmetry between F#-F and F-F# (because G is higher in the tonal hierarchy than F).

These principles were proposed as statements of the psychological effects of the tonal hierarchy independent of the particular experimental measure used, which might be direct judgments, memory accuracy, event-related potentials, or other measures.

3.4 Contemporary Issues that Arise from These Basic Studies

These basic studies have raised a number of issues that are considered next. One issue is whether and how tonal hierarchies are learned. A second is the question of whether tonal hierarchies are musical facts, that is, can be related to objective properties of the music itself. A third is how computational models might serve to understand the structure and origin of tonal hierarchies and how they might use tonal hierarchies to model perceptual processing of music. Finally, we consider the role that tonal hierarchies have played in a recent music theoretic proposal in which it is used to compute distances between musical events and make testable quantitative predictions.

Concerning the first issue, a specific learning-based proposal is that tonal hierarchies require extensive experience with music to be internalized. Through repeated and extensive exposure, listeners have learned the relative positions of tones in the tonal hierarchy. Another learning-based approach suggests that learning occurs over a much shorter term. Tonal hierarchies may result from actively processing the musical input forming summaries of statistically frequent tones and tone

combinations. If so, then psychological measurements, such as probe tone ratings, may reflect short-term memory for the preceding context. A third nonlearning, psychoacoustic explanation is that the tonal hierarchy reflects acoustic properties of tones. These depend on the harmonics of complex tones in a way that is described later.

To assess these alternatives, a variety of approaches have been taken. Some studies use development and music training as a way to determine the importance of experience on acquiring the tonal hierarchy. Another empirical approach examines individual differences and neurological case studies for abilities allied with the recovery of the tonal hierarchy. This may give clues as to the processes through which tonal hierarchies are acquired. In addition to these commonly used approaches in psychology, music offers another alternative, which is to employ unfamiliar musical styles, for example, from other cultures or nontonal Western music. Another approach is to develop computational models to simulate the empirical results. This approach has the potential for identifying musical features important to establishing tonal hierarchies, and may suggest processes through which they may be established cognitively.

3.4.1 Developmental Studies

The learning-based accounts just outlined assume the tonal hierarchy is internalized through exposure to music. This proposed learning mechanism makes a specific prediction about the developmental course of acquisition. If tonal hierarchies are implicitly acquired through exposure to music, then apprehension and representation of tonal hierarchies will emerge at a later developmental stage than the basic perceptual sensitivities on which they are built. The reason is that if tonal hierarchies are to become internalized as cognitive resources, they require a mature memory system and specialized interactions with environmental resources.

This important statement is associated with a basic conundrum. Because statistical learning may occur in infancy, why does the acquisition of the tonal hierarchy, assumed to be the result of statistical learning, occur relatively late? The problem may be resolved by proposing that although the infant brain has developed to the extent of extracting simple regularities in sound patterns, it has not yet developed those memory resources that along with musical experience allow the extraction of hierarchical regularities among tones.

Numerous studies of infant and child development support this proposal. Regarding basic perceptual sensitivities, during the first year of life infants develop an impressive repertoire; they develop the discrimination of, for example, melodic contours, frequencies, simple harmonic ratios, phrasing, and, to some extent, pitch-scale patterns (for reviews tracing this first year, see Trehub and Trainor 1993; Trehub et al. 1997; Dowling 1999; Cohen 2000; Trehub 2000). However, regarding apprehension and representation of tonal hierarchies, evidence for appreciation of Western tonal structure does not appear for several more years.

The appearance of a stable tonal center appears to emerge around 5 or 6 years of age. This is the first age at which a stable tonal center is evident in children's spontaneous singing (Dowling 1999). Similarly, Zenatti (1993) reported a preference for tonal over atonal contexts at 5 or 6 years (depending on the musical task); the distinction tended to increase over 8–10 years for most children. The ability to process a tonal melody is suggestive of the internalization of a tonal hierarchy that guides encoding and retrieval of melody tones. Trainor and Trehub (1994) evaluated the ability of children and adults to detect changes in a well-structured Western tonal melody. The 5-year-olds in their study were able to detect changes that were out of key but not changes that altered the implied harmony. The 7-year olds and adults noticed changes both to the key and the implied harmony.

An early probe-tone study by Krumhansl and Keil (1982) asked children in first through sixth grade and adult listeners to judge the “goodness” of a six-tone melodic pattern. The initial four tones of the pattern were the major triad tones C-E-C-G presented melodically. The final two tones were probe tones located one octave distant from the triad. The judgments of the youngest children showed only a distinction between scale and nonscale tones. Older children distinguished triad tones from nontriad tones; only the adults isolated the tonic from the other tones of the triad, and revealed the full tonal hierarchy in their judgments. It was noted that for all listeners the scale-step distance between the two probe tones also influenced judgments.

The children's task was simplified in another study (Cuddy and Badertscher 1987). They asked children in the first to sixth grades to rate one probe tone following the C-E-C-G pattern for “goodness” of completion of the pattern. The pattern and probe tones were Shepard tones (Shepard 1964) intended to reduce attention to pitch distance and focus attention on tonal relations. Under these conditions, Cuddy and Badertscher found that even the youngest children differentiated the levels of the tonal hierarchy – the tonic, the other triad tones, the scale tones, and the nonscale tones. Along with Speer and Meeks (1985), this study also found that major scale patterns were also effective for yielding evidence of the tonal hierarchy.

Subsequent developmental data suggested a subtler trend in the acquisition of knowledge of tone relationships (Lamont and Cross 1994). Collected with two variants of the probe-tone tasks and a game-playing task, their data from a large sample size ($N = 285$) revealed understanding of scale structure in the youngest children with increasing sophistication up to 11 years of age. Thus, the estimate for acquisition of the tonal hierarchy appears to vary with task demands and strategy, but preschool or early school years are clearly important benchmarks (Shuter-Dyson 1999).

The assumption that tonal hierarchies are implicitly acquired through exposure to music yields the corollary that their apprehension does not require formal music training. Such a claim may seem surprising on first consideration, given the large amount of evidence indicating the advantage of music training for multiple musical tasks. However, the acquisition of the tonal hierarchy may have a privileged status.

That young children reveal this form of tonal knowledge rules out an account based on skills acquired during the study of advanced music theory or required for expert performance.

Thus, for adult listeners, no role of early music training in their representation of the tonal hierarchy would be expected. Here current experimental evidence is somewhat mixed. Some evidence suggests that music training does enhance performance on tests of tonality (Krumhansl and Shepard 1979; Jordan and Shepard, 1987; Frankland and Cohen 1990; Steinke et al. 1997). Other studies of the tonal hierarchy, however, tend to support the claim that music training may not be required. Relations between probe-tone results and music training were either not statistically significant for the melodic pattern C-E-C-G (Cuddy and Badertscher 1987; Brown et al. 1994) or were slight (Cuddy 2000). Minor differences in procedure, such as the choice of pure versus Shepard tones to construct probe-tone stimuli, may be partially responsible for the differences in findings. Untrained listeners tend to focus on pitch height in probe tone responses; as noted earlier, Shepard tones deemphasize pitch height and may help direct the untrained listener to focus on tonal relations. Another difference is sample size, with large samples ($N = 100$) yielding more statistical power to pick up slight differences (Steinke et al. 1997). Yet even that study showed that the response patterns for all levels of music training were essentially similar to music theoretic descriptions of the tonal hierarchy. The more musically trained simply showed finer distinctions among levels of the hierarchy.

Other paradigms implicating the tonal hierarchy point to the similarity of responses for trained and untrained listeners. In tests primarily designed to assess Narmour's (1990) bottom-up principles of melodic expectancy, the influence of tonal hierarchy on expectancy judgments was found for both trained and untrained listeners (Cuddy and Lunney 1995; Krumhansl 1995a, b; Thompson et al. 1997). Finally, although music training was associated with improved recognition memory for unfamiliar tunes (Cuddy et al. 1981, 2005), both musically trained and untrained listeners responded similarly to the tonal structure of the tunes.

Taken together, the evidence suggests that a representation of the tonal hierarchy exploits general perceptual predispositions, is acquired in childhood through acculturation, and is acquired without formal intervention. Tonal melodies are easier to recognize than nontonal melodies; pitch alterations in melodies are easier to detect if the alteration deviates from tonal rules than if it does not. Music training is not needed to facilitate acquisition of the tonal hierarchy but, rather, teaches skills and strategies to apply this knowledge to musical problems. (For supporting arguments, see Bigand and Poulin-Charronnat 2006; Marmel and Tillmann 2009.)

Moreover, this section has elaborated an application of the second basic cognitive principle underlying tonal hierarchies, that of sensitivity to statistical regularities in music. This sensitivity is not dependent on music training, but is dependent on the maturation of a memory system capable of dealing with tonal/hierarchical regularities. Thus musical memory and representation of the tonal hierarchy are intimately associated. Across individuals, covariation between these two components of musical cognition may be expected.

3.4.2 *Individual Differences and Neurological Case Studies*

The argument is that the tonal hierarchy provides a stable framework for accurately encoding and remembering musical pitch patterns. From this argument it follows that the cognitive representation of the tonal hierarchy should be associated with musical memory. One test of this statement is to examine individual differences in associations between tonal hierarchy and musical memory. Individuals differ in perceptual and cognitive abilities and also in early musical environments. Thus, individual differences should be revealed as follows: the ability to recover the tonal hierarchy is reflected in proficiency at musical memory; failure or loss of the tonal hierarchy is accompanied by musical memory failures.

Striking observations of individual differences result from the comparison of neurologically disordered individuals and healthy age-matched controls (see, e.g., the special April 2008 issue of *Music Perception* which is devoted to Music and Neurological Disorders). Neurological case studies examine patterns of loss and sparing of abilities, called “dissociations.” One line of evidence has been obtained from case studies of individuals with localized brain injury resulting in selective loss of musical abilities (for reviews, see Peretz 1993a; Dalla Bella and Peretz 1999; Marin and Perry 1999).

Interpreting the early historical evidence is a challenge. The evidence is anecdotal, not collected under controlled conditions. Individual patients may display complex, not regular, patterns of symptoms and many questions regarding cerebral localization of musical function remain unanswered. Nevertheless, recent case studies of amusia – clinical disorders of music abilities due to brain damage – have reported data collected under controlled conditions and reveal much about the functional architecture of the brain.

Peretz and colleagues (e.g., Peretz 1993b, 1996; Peretz et al. 1994, 1997; Liégeois-Chauvel et al. 1998) have systematically investigated three patients, CN, GL, and IR, who sustained musical deficits after surgery involving bilateral damage to auditory cortex. An extensive program of testing included music, speech, language, and other cognitive skills. Case results were compared with results from controls matched for age, education, and music background (all were nonmusicians). A fourth patient, KB, an amateur musician who suffered right frontoparietal damage after a stroke, was similarly evaluated by Steinke et al. (2001; see also Lantz et al. 2003).

All four patients had normal language and intellectual functioning and all four failed various tests of musical functioning. Thus, the results clearly demonstrate dissociation (separation) of music and language in brain organization. Within the music tests, the patients had difficulty with both tests of tonality and tests of recognition memory for familiar music (for the latter, KB’s difficulties extended to instrumental, but not song, tune recognition). The connection between representation of the tonal hierarchy and musical memory is supported most clearly in the data of CN and GL. These two patients, unlike IR and KB, showed good recovery (CN) and sparing (GL) of perceptual skills. Thus, failures at the tests of tonality and memory are not likely due to impaired perception of tones.

For example, 6 years post-onset CN performed at a normal level on auditory perceptual tasks, such as discrimination of isolated musical pitches, detection of contour and interval (frequency ratio) changes in novel musical sequences, and detection of rhythmic changes in novel musical sequences (Peretz 1996). However, CN was still markedly impaired at recognizing and naming familiar melodies, and classifying melodies as familiar or unfamiliar. She was also impaired at rating probe tones with a tonal melody as musical context, at judging the appropriate tonal ending for melodies, and at remembering the pitches of unrelated tones (Steinke et al. 1994, 1997). Given a probe-tone task (that of Cuddy and Badertscher 1987), GL's responses showed some evidence of contour and interval processing, but no sensitivity to tonal function (Peretz 1993a, b). GL was also deficient at melody recognition and pitch memory. Peretz (1996) suggested that tonal encoding of pitch (coding in terms of tonal function) is a major determinant of access to stored musical representations. From a review of the evidence, it is probably critical. Impairment of acquired cognitive references and/or their implementation leads to difficulties with the probe-tone and other tonality tests and severe failures of melody recognition and memory.

Complementary evidence also supports a link between tonal encoding and musical memory. This evidence has been obtained in case studies of (probable) Alzheimer's disease (AD) – for example, Cuddy and Duffin (2005), Fornazzari et al. (2006), and Vanstone and Cuddy (2010). Rather than displaying musical impairments after stroke despite cognitive recovery, some AD individuals show the reverse pattern of dissociation. They demonstrate preserved musical memory despite severe speech and other cognitive difficulties. Of importance, in the studies by Cuddy and colleagues, is that these persons detected tonal errors in tunes as accurately as did age-matched healthy controls. Thus intact tonal encoding of pitch may have facilitated access to stored representation of tunes.

As a final example under the topic of individual differences, cases of congenital amusia may be noted. Unlike brain damage, congenital amusia (or “tone deafness”) is not acquired through injury or assault. It is considered a developmental disorder (e.g., Ayotte et al. 2002; Hyde and Peretz 2004; Peretz et al. 2008) and may be a neurogenetic anomaly (Drayna et al. 2001; Peretz 2008). The auditory skills of congenital amusics are neurologically normal, with the outstanding exception of the inability to develop normal musical abilities such as recognizing a familiar tune in the absence of lyrics and detecting out-of-key notes in conventional melodies. They lack “the (implicit) knowledge and procedures required for mapping pitches onto musical scales” (Peretz et al. 2008, p. 332).

Thus, in certain cases of acquired and congenital amusia, both musical memory and tonal representations are compromised. Musical memory, it has been suggested, depends on top-down activation of a cognitive framework for encoding and storing musical materials. As revealed by studies of individual differences and neurological case studies, the tonal hierarchy is a crucial component of this framework.

3.4.3 *Tonal Hierarchies and Tone Distributions in Western Music*

If listeners acquire the tonal hierarchy by internalizing statistical regularities on the musical surface, then psychological measures of tonal hierarchies should correlate with tone distributions in musical compositions. This section reviews a number of studies that find the tonal hierarchy measured in empirical research mirrors the emphasis given the tones in compositions by frequency of occurrence and duration. This relationship between subjective and objective properties of music provides a strong musical foundation for the psychological construct of the tonal hierarchy.

According to Meyer (1956), musical styles consist of systems of probability relationships that capture the characteristic patterns in the style. He goes on to say that the meaning of any given musical event depends on how it enters into the patterns captured in the probabilities. Thus, the hierarchy evident in the probe-tone ratings, and the variety of other behavioral measures summarized earlier, might be related to the statistical distributions of tones in music.

This relationship was first examined by Krumhansl (1985; see also Krumhansl, 1990a), who compared the tonal hierarchies with various statistical treatments (Youngblood 1958; Hughes 1977; Knopoff and Hutchinson 1983). These statistical treatments tabulated the frequency of occurrence or total duration of each tone of the chromatic scale in pieces by Schubert, Mozart, Hasse, Strauss, Mendelssohn, and Schumann. Tone distributions correlated strongly with the probe-tone profile for the corresponding key; tones high in the tonal hierarchy tend to be sounded more frequently and for longer durations. Similar results were found by Järvinen (1995) in jazz improvisations, especially at more stressed metrical positions. Minor discrepancies between the statistics and the probe-tone profiles have been noted (Krumhansl 1990a, p. 69; Auhagen and Vos 2000) but these were relatively minor and can be explained in large part by proximity to the tonic.

Huron has been at the forefront of analyzing music for statistical properties (see the summary in Huron 2006). He has also developed and distributed tools for statistical studies of music (Humdrum Toolkit and Themefinder.com). He stresses the adaptive value in evolution of being able to anticipate frequently occurring events (Huron 2006, p. 357). As a general cognitive principle it is adaptive because knowing what, when, and where something is likely to occur speeds perception, action, and evaluating the consequence of alternative actions. He also notes the advantages of statistical learning, “In a stable environment, the most frequently occurring events of the past are the most likely events to occur in the future. Thus, a simple yet optimum interactive strategy is to *expect the most frequent past event*.” (p. 360, italics in the original)

In an early study, Huron (1993) analyzed a large sample of Bach’s music. He looked at which tones in chords are most frequently sounded in two different octaves (doubled). The number of these doubled tones correlates strongly with the

probe-tone profiles. He concluded that the doubling of the tones high in the tonal hierarchy reinforces the perception of key. Huron's collaborator Aarden (2003) conducted the largest note count to date. As shown in Huron (2006, pp. 148–149), Aarden found the distribution of tones in music (when modulating passages were excluded), with more than 65,000 notes for melodies in major keys and more than 25,000 notes for melodies in minor keys. Aarden's note counts differ somewhat from previous ones. This may reflect the different compositions analyzed, but Aarden went on to show that listeners' expectations conformed better to his statistical results than to previous ones. Thus, the behavioral measures were accounted for better by his musical analysis of tone distributions than by other analyses.

Although the details of the tone distributions in these studies differ in detail, they all point to a common conclusion. Subjective properties of music as studied in psychological research correlate well with objective properties of music, specifically the relative emphasis to the tones by frequency and duration. The distributional emphasis would serve the purpose of initially establishing and then maintaining the listener's sense of the tonal reference points. The extent to which the probe-tone ratings resemble the tone distributions supports Meyer's proposal that listeners have internalized the statistical properties of music.

3.4.4 *Cross-Cultural Studies of Tonal Hierarchies*

One way to study the acquisition of tonal hierarchies is to use unfamiliar styles from non-Western cultures. It is important to look at tonal hierarchies cross-culturally to explore their generality across styles. Moreover, it is important to rule out competing explanations of their psychological basis. That is, the results just discussed for Western music may be a function of some other musical attribute. For example, perhaps consonance influences both the distribution of tones and the measured tonal hierarchy. More specifically, Western music is organized around harmony (chords) and so it might favor tones that form consonant intervals with the tonic and other tones that are consonant with the tonic, and this might determine the tonal hierarchy. This can be addressed by studying music that is not organized around Western harmonic structure.

Two studies bear directly on the acquisition of tonal hierarchies. These collected probe-tone ratings with contexts drawn from non-Western music. Castellano et al. (1984) used contexts from 10 North Indian *ragas*. One of the most significant differences from Western music is that the primary means of expressing tonality in Indian music is through melody. In addition, North Indian music has a greatly expanded set of scales (called *thats*, all with the same tonic) compared to the major/minor system of Western music. Theoretical treatments of Indian music describe a hierarchy of the importance of tones. In a probe tone study, ratings of the 12 tones of the Indian scale largely confirmed these predictions.

It was surprising that the results for both Western and Indian listeners agreed with the theoretical predictions for the Indian music. This was surprising because

it indicates that the Western listeners adapted quite readily to tonal hierarchies in this unfamiliar style, rather than requiring extensive experience. Both groups gave high ratings to the first (*Sa*) and fifth (*Pa*) scale tones, which are described as most structurally significant and which are sounded in the drone accompanying the melody. They also gave relatively high ratings to the *vadi* tone, which is a tone given emphasis in the melody and is specific to each *raga*. To explain the agreement between the two groups of listeners, the contexts were examined. It was found that the theoretically important tones were sounded more frequently than other tones. Apparently, the unfamiliar listeners used this information about the distributions of tones to make their judgments. It suggests that distributions of tones can convey the tonal hierarchy to listeners unfamiliar with the style. Beyond this level of agreement, the Indian listeners were more sensitive to scale (*that*) membership. The important finding, however, is that the results demonstrate a strong link between objective and subjective measures of tonal hierarchies.

Similar results were found in an experiment with the music of Bali (Kessler et al. 1984). The study of Kessler et al. had an expanded design compared to the study with North Indian music. It used contexts from both Western and Balinese melodies, and both Western and Balinese listeners some of whom were unfamiliar with Western music. Balinese (and Javanese) music is interesting because it uses two different tuning systems, *sléndro* and *pélog*, both of which are different from Western (diatonic or chromatic) tuning (see also Perlman and Krumhansl 1996). Their study included both *sléndro* and *pélog* contexts. Listeners' responses revealed a number of strategies. Some, but only some, of the Balinese subjects produced results corresponding to the predicted tonal hierarchies. Other Balinese subjects responded primarily to pitch height, giving higher ratings to higher tones. Western listeners responded to frequency of occurrence of the tones in the context, which also correlated to some degree with the predicted hierarchy. Again, this suggests that naïve listeners can use tone distributions to perceive the major anchoring tones of the tonality.

To our knowledge these two studies are the most directly related to tonal hierarchies in particular. Two other studies demonstrate in cross-cultural studies that the tonal hierarchy is one factor entering into judgments of melodic continuations (Krumhansl et al. 1999, 2000). The first of these used Finnish folk hymns that are passed on by oral tradition by conservative religious sects in southwest Finland. They combine elements of Finnish folk music and Lutheran hymns as they came to be known in Finland in the early eighteenth century. They contain modes and other features unlike tonal-harmonic music. The second study used Sami *yoiks*, which are also purely oral and improvised around short, repeated motives. Most are based on the five-tone pentatonic scale and the intervals in the *yoiks* tend to be larger than those found in most Western melodies. Listeners unfamiliar with the style demonstrated that they were able to orient to the appropriate tonal hierarchy based on the relative emphasis given the tones in the music. As in other studies, the melodic continuation judgments of listeners familiar with the style exhibited style knowledge over and above the salience of tones in the context.

Recent research in music cognition has expanded to include styles from many different cultures, studying not only the way in which the music is perceived and understood, but also its emotional effects. Cross-cultural studies raise a vast number of interesting questions and may reveal underlying cognitive principles not yet identified in music research.

3.4.5 *Tonal Hierarchies in Nontonal Western Music*

More recent Western music has introduced styles of music that depart from traditional tonal structure. In other words, they are not structured around the system of major and minor harmony that dominates Western music. These novel styles raise the question as to whether the structural principles with which they are written can be perceived and understood by listeners. Twelve-tone serialism has been particularly controversial. In addition to investigating the listener's response to the styles generally, these studies provide additional materials with which tonal hierarchies can be examined.

Two studies used the probe-tone method with bitonal contexts, in which tones from two different keys are played simultaneously. The style raises the question as to whether listeners can process two keys as separate entities or whether the resulting perception is one of a fused combination of the two. In the first of these (Krumhansl and Schmuckler 1986), the context was the *Petroushka chord* from Stravinsky's ballet. It is a striking example of bitonality that uses the tonic triads of the keys of C and F# major. These keys have maximally dissimilar tonal hierarchies (Krumhansl and Kessler 1982). This would seem to optimize the possibility that the materials from the two keys could be heard independently, as there would be little confusion as to which key a tone belonged. Indeed, the probe tone ratings showed the influence of the two keys, but two experiments with selective attention tasks (wherein participants were instructed to focus on just one key) established that listeners cannot perceptually segregate the materials from the two keys. This was true even for a group of musicians who had recently performed the piece in concert, including the two clarinetists playing the part. These findings suggest that the two component keys are not perceptually functional as independent entities. Rather, the context appears to establish a composite hierarchy of the tones of the two triads.

The second study of bitonality by Thompson and Mor (1992) found more positive evidence for listeners' abilities to process two keys at once. Their two excerpts (from pieces by Dubois and Milhaud) contained materials from the keys of C# and F major, one represented in the upper stave and the other in the lower. In both, the materials in the two staves were quite distinct. In the first, the music in the upper stave was melodic whereas that in the lower stave was chordal. In the second, the music in the upper stave was considerably higher in pitch than that in the lower stave. With these differences, the listeners' responses reflected long-term knowledge of tonal hierarchies, not just the distribution of tones in the contexts. In the

case of the first excerpt, evidence for two functional keys was found; in the case of the second excerpt, the tonality of the material in the upper voice strongly predominated.

Krumhansl et al. (1987) examined another unconventional style, 12-tone serialism. Twelve-tone serialism is an influential style in which many compositions were written in the twentieth century. It is of special interest here because it is specifically intended to oppose traditional ways of structuring music in terms of chords and keys, not allowing a hierarchy of salience to emerge. It does this by requiring that all tones of the chromatic scale be sounded before any of them is repeated. The ordered series of the 12 chromatic tones is called the *series*, or *tone row*. This means that no tone receives particular emphasis. This study differed from those reviewed in the last section because it found different results depending on familiarity with the style.

The experiments used materials from two compositions by Schoenberg (*Wind Quintet and String Quartet*, No. 4), who is generally considered the innovator of the style. Two groups of listeners participated, both of which were musically trained. However, only one group was familiar with the 12-tone style primarily through academic study. Their probe tone ratings showed evidence of reversing the normal pattern for major and minor tone profiles – that is, they gave high ratings to tones that denied local implications of key. This is consistent with the intention of the style to avoid tonal implications. In contrast, the probe tone ratings of the group of listeners unfamiliar with the style showed influences of local tonal implications, contrary to the style's intention, and also gave higher ratings to tones sounded more frequently, as in other studies reviewed here, and also more recently, which would be expected from short-term memory. Parenthetically, neither group of listeners seemed to have internalized the ordered sequence of tones in the series. They did not give higher ratings to tones following the contexts in the series, despite rather extensive experience with the series in the course of the experiments.

In sum, these studies of nontonal Western music show that the influences of major and minor tonal hierarchies are more complex than in traditional Western music. The experimental results are in some cases consistent with the compositional methods, but in other cases they are not. Given the vast number of recent stylistic innovations, these studies can be viewed only as a minor inroad into the interesting questions they pose.

3.4.6 *Novel Tone Sets*

The work cited in the preceding text indicates that the distribution of tones in composed music supports the tonal hierarchy. Frequencies of tone occurrence and tone duration both tend to correspond with position in the tonal hierarchy. According to Krumhansl (1990b), “the primary significance of the observed correspondence is to suggest a mechanism through which the principles of musical organization are learned” (p. 315). With repeated listening to music, the consistent features of the

tone distribution become internalized as an abstract internal schema or framework (Krumhansl 1987, 1990a).

The question then arises as to whether the internal schema, thus acquired, interferes with or even cancels the listener's ability to pick up pitch distributional information in a novel idiom – that is, a distribution not convergent with the Western tonal hierarchy. If the pitch distribution information in a musical context is not convergent with the tonal hierarchy, one of several possibilities may result. The listener may be unable to pick up or remember conflicting information; the information is either not processed or not retained. Alternatively, a listener may attempt to assimilate the conflicting information to a more familiar tonal hierarchy (Dowling 1978). Perceptual judgments may be systematically distorted with respect to the pitch distribution (Jordan and Shepard 1987). Or, finally, listeners' strategies for the abstraction of pitch structure may be flexible and adaptable. Thus, they may easily abstract distribution information that deviates from the conventions of the Western tonal idiom (Krumhansl 1990a).

Oram and Cuddy (1995) addressed the question in the following way. They constructed melodic sequences of 20 pure tones, each tone of duration 200 ms. Each sequence was generated from either a diatonic tone set, C, D, E, F, G, A, B, or a nondiatonic tone set. If nondiatonic, the tone set did not conform to the scale of any major or minor key. Within each sequence, the frequency of occurrence of each tone was determined according to the following ratios: one tone of the tone set occurred eight times, two tones occurred four times each, and the remaining four tones occurred just once. The tones selected to occur most frequently never formed the simple pattern of a major triad in root position. Both musically trained and untrained listeners heard the sequences in a probe-tone paradigm. The results for both levels of music training were quite straightforward. Probe tone ratings were systematically related to the frequency of occurrence of the tones in the sequence.

Musical knowledge, however, did play a role. First, the effect of frequency of occurrence within the melodic sequence was more pronounced for musically trained than untrained listeners. This finding is surprising and somewhat counterintuitive. Yet it is both consistent with the notion of flexibility of musical pitch processing and suggestive that music training does not tie processing to a rigid schema – rather it fosters a strategy for the pick-up of novel distributional information.

Second, the effect of frequency of occurrence was more pronounced for diatonic sequences than for nondiatonic sequences. This indicates that prior knowledge of diatonic structure can be coordinated with sensitivity to tone distributions. Third, the data for the musically trained listeners revealed some degree of assimilation to the Western tonal hierarchy for diatonic sequences. Along with frequency of occurrence and pitch proximity (pitch distance of the probe tone from the last tone of the sequence), relative stability in the tonal hierarchy added to prediction of the probe-tone ratings. The important conclusion from this study, however, is that listeners are flexible and adaptable to novel pitch distributions, and that extensive music training does not appear to interfere with this sensitivity – if anything, it appears to enhance it.

Cuddy (1997) reported a follow-up probe-tone study with composed musical melodies. As before, sequences were generated from either a seven-tone diatonic or

nondiatonic tone set, but the sequences were now 20-s original flute compositions. The total duration of one tone, summing across each occurrence of the tone, was 8 s; of two different tones, 4 s, and of four different tones, 1 s. Within the total duration allotted to each tone, composers were allowed to distribute the duration of each occurrence of that tone and its octave location according to their own intentions or style. Despite differences in musical materials, the results closely replicated those of Oram and Cuddy (1995). Durational differences in the musical surface were reflected in probe tone ratings.

Cuddy (1997) also reported data from a second study in which a pair of probe tones followed each presentation of the flute melody. The listener was asked to judge how related the first tone was to the second tone of the pair with respect to the melody. For Western tonal harmonic contexts, relatedness is judged by order of the tonal stability of the two probes, as noted earlier in this chapter (Krumhansl 1979; 1990a; Bharucha and Krumhansl 1983). With the novel flute melodies, on the other hand, judgments of relatedness were predictable not from tonal stability, but from the duration biases of the melody. They were higher for the order in which a probe tone of shorter total duration in the melody was followed by a probe tone of longer duration in the melody, than the reverse.

This section proposed that the correspondence between distributional properties in music and the tonal hierarchy is important because it suggests a mechanism for how the tonal hierarchy is acquired. Work with novel tone sets reveals that listeners do possess a finely tuned ability to discover distributional regularities, an ability that is a necessary prerequisite to learning. They also preserve the ability to organize tones in new musical contexts.

3.4.7 Tone Distributions: Frequency or Duration?

A question that arises from these studies is whether pitch salience depends on the frequency or duration of the tones in the inducing context. The paradigm with novel tone sets allows researchers to tease apart the role of distributional cues that are tightly coexistent in music. The mechanisms for the cognitive processing of frequency of occurrence – as reflected by frequency of tone onset – may differ from those for processing tone duration. The (controversial) argument is that the two processes are separable both at the neural level (Whitfield and Evans 1965) and at the general level of cognitive principles (Yonelinas et al. 1992). It is thus instructive to consider whether separation is applicable to musical contexts. If it is, implications follow for the relative importance of cues in key-finding both by listeners and by computational models (see the next sections).

Salience was manipulated in Oram and Cuddy (1995) as the frequency of occurrence of isochronous tones; thus frequency of occurrence and total duration covaried. Salience in Cuddy (1997) was defined as total duration of tones with frequency and duration of individual events free to vary. Nevertheless, the findings were the same: Tones of greater surface salience in the sequence, not the diatonic tonal hierarchy,

acted as reference points or anchors for other tones of lesser surface salience. Of the possible outcomes outlined at the beginning of this section, the one favored is that listeners' strategies for the abstraction of pitch structure were flexible and adaptable. Representation of the Western tonal hierarchy did not interfere with the abstraction of distribution information that deviated from familiar convention.

Lantz and Cuddy (1998) constructed nontonal melodic sequences in which frequency of occurrence and duration were pitted against each other. For example, in a given sequence, certain tones occurred more frequently, while the duration (note values) of other tones was longer. In addition, the total duration of the longer tones was varied. Using a standard probe-tone paradigm, the researchers found that tones sounded with longer duration, and tones sounded for greater total duration, were rated higher than other tones. Frequency of occurrence per se did not influence listeners' ratings under these conditions.

In related work, Smith and Schmuckler (2004) also used the probe-tone technique to examine cues of frequency of occurrence and duration. Their technique of sequence construction was slightly different from those of the aforementioned studies. Their sequences contained all 12 tones of the chromatic scale presented in random order. The assignment of frequency of occurrence and duration to each of the 12 tones was varied across experiments. Moreover, the duration of each tone was assigned to correspond or not to correspond with the salience of each tone in a given key of the Western tonal hierarchy. Convergent with the aforementioned studies, Smith and Schmuckler (2004) found that duration was a more salient cue to structure than frequency of occurrence. They also noted that duration was increasingly effective as the absolute difference increased between the duration values assigned to tones. However, their listeners were sensitive to duration only when both the individual and total duration of tones corresponded to the Western tonal hierarchy. In other words, listeners' ratings did not differentiate reliably among tones if the assignment of duration values did not correspond to this hierarchy. This work, therefore, suggests some limitations on listeners' sensitivities to novel pitch distributions. One possible factor playing a role in setting limitations is the overall pitch information load. Listeners' capacities for apprehending novel pitch distributions through duration cues may be limited to tone sets of six or seven tones (as is found in most musical scales) and may not extend to tone sets of 12 tones.

In sum, research with novel tone sets promises to uncover sensitivity to the cues leading to acquisition of the tonal hierarchy and also sensitivity to cues pointing to nontraditional or nontonal hierarchies. More research is invited to clarify the extent and boundaries of these sensitivities.

3.4.8 Tonal Hierarchies and Computational Models

Many computational models of tonality have been proposed and tested against psychological data. Some of them focus specifically on the tonal hierarchy. One reason for this is that the quantified tonal hierarchy provides detailed data against

which the computational models can be tested. Different models make different assumptions about the processes and levels of processing at which the tonal hierarchy is generated and the way it is manifested in the output of the model.

One model uses psychoacoustics, that is, low-level processes to generate tonal hierarchies. This model tests the extent to which psychoacoustic processes can account for tonal hierarchies without assuming cognitive processes. At the core of Parncutt's model (1989; 1994; Huron and Parncutt 1993) is the notion of pitch salience, which derives from Terhardt's (1979; Terhardt et al. 1982a, b) model of pitch perception. This approach posits subsidiary pitches that are not physically present, but that arise through the interaction of frequencies that are present. These are called *virtual pitches*. The virtual pitches are weighted according to how well the spectral components match its harmonic series or harmonic "template." In one application, Parncutt (1994) found that this approach could model quite well the tonal hierarchies of some of the context sequences used in the first experiment of Krumhansl and Kessler (1982). The correlations with the probe-tone results were quite high. To eliminate the possibility that the probe-tone profiles simply reflect the tone distributions in the experimental contexts, they also compared the results of their model with the distributions of tones in the contexts. That these correlations were lower argues against the idea that the probe-tone profiles simply reflect tone distributions in the contexts, and suggests that the virtual pitch approach may account for some of the patterns in the probe-tone profiles that go beyond the presence or absence of tones. In an extension of this model, Huron and Parncutt (1993) combined the pitch salience approach with sensory memory decay and again found fairly close results.

Another model uses a multilevel approach, assuming processes at both psychoacoustic and cognitive levels. Leman's (2000) model incorporates both virtual pitches, as in Parncutt's model, and short-term memory. The input to the model is acoustic, which is processed by a peripheral auditory model (simulating the filtering of the ear), and then analyzed for periodicity pitches. The summed "completion image" is similar to the virtual pitch model just described. This pitch module is then entered into an echoic memory module, which incorporates both integration and decay over time. Applied to the chord contexts of Krumhansl and Kessler (1982), the model produced tonal hierarchies that are similar to the probe-tone profiles. From this, Leman concluded that the probe-tone profiles could be accounted for by short-term memory for the perceptually immediate context. There are a number of problems with this conclusion, the most decisive of which is that his input was the composite of all harmonic contexts in Krumhansl and Kessler (1982), not the individual harmonic contexts. Direct comparisons would show marked discrepancies. Moreover, this model's result for scale contexts primarily reflects scale membership and not the probe-tone data.

A different, more dynamic, approach involves self-organizing neural networks. Such a model has been presented by Tillmann et al. (2000). A self-organizing map creates a topographic mapping between the input data and neural net units on the map. Before learning, the map units have no particular organization, but when presented with input data an ordering appears over time. The model contains three layers,

corresponding to tones (pitch classes), chords, and keys (only major keys were represented). Different training sets were used: simple (idealized) harmonic sequences and more realistic harmonic sequences. These were either “sparsely coded” (just the presence of the tones was coded) or “richly coded” (including psychoacoustic pitch salience following the scheme of Parncutt 1988). In other words, one model assumes that the pitches are simply perceived as such, and the other model takes into account psychoacoustic properties.

After training, the neural network was tested against various previous experimental studies. In one test, their trained model was “presented” with major and minor chords. The resulting activation of tone units resembled the probe-tone profiles of Krumhansl and Kessler’s (1982) first experiment. Their model was also compared with the results for the three major key sequences of Krumhansl and Kessler’s (1982) second experiment (the model only included major keys). In that study, sequences of chords were played stopping at each successive chord for probe tone judgments in order to trace how the sense of key developed and changed over time. Their model traced quite closely these changes in the sense of key. Finally, the neural network model reproduced the direction and magnitude of similarity judgments for sequentially presented tones. (Recall that two tones are judged to be more related if the second tone is higher in the tonal hierarchy than the first, compared with the opposite order; Krumhansl 1979, 1990a).

Together, the models described in this section offer suggestions about possible psychological mechanisms that contribute to the perception of the tonal hierarchy. They variously include virtual pitches created by interactions of physically present harmonics, memory traces of the perceptually immediate context, and the result of internalizing regularities from harmonic sequences over an extended period of time. At present, no single account seems entirely adequate to account for the wide range of behavioral data currently available on tonal hierarchies. It seems that the most likely outcome of further modeling efforts will be that the behavioral data rely to some degree on all three of these psychological mechanisms, as well as others, and that these various mechanisms are not clearly independent of one another.

3.4.9 Using Tonal Hierarchies in Key-Finding Models

A number of models have addressed the problem of key-finding. Key-finding refers to the process through which a listener initially orients to the key of a piece of music and subsequently reorients to new keys if modulations (changes of key) occur in the music. It is important for music perception because the function of tones in melodies and chords depends on their relationship to the tonic, as described previously.

The objective of key-finding models is to take some musical input and assign a key (or keys) to it. Models vary as to whether the input is acoustic information or symbolically coded music (such as score notation or MIDI code). The choice of input also depends on the objective of the model; it may be an entire piece of music,

a short segment from the initial or final portion, or a sliding window of some length. Some models allow retrospective reevaluation of key assignment. Some models assign a single key and other models allow for the possibility that a number of keys might be quite strongly suggested (or no key may be perceived as very strong). Key-finding models may or may not attempt to characterize the process through which the listener finds the key, however. That is, some attempt to be psychologically realistic, others draw on music theory, and others are shaped by computational considerations. Whether intended as a psychological model or not, automatic determination of key has utility in applications, for example, determining the key is prerequisite to successful automation of music analysis (Rowe 2000). Considerable modeling effort has been devoted to the problem of how to characterize this process (cf. Vos and Leman 2000), and only those that use tonal hierarchies are described here.

Likely the first key-finding algorithm to be implemented on a computer was that of Longuet-Higgins and Steedman (1971). This algorithm uses a two-dimensional array of tones. It matches incoming tones of a piece of music to box-shaped regions containing the scale tones of a key. In their model there is one region for each major and minor key (the harmonic minor scale is used). The algorithm works by eliminating musical keys as the music progresses. Special rules are applied if the model finds either no or multiple keys at the end of the process. Its results can be compared with models that assume that over and above scale membership the tones are differentiated more finely, as in the tonal hierarchies.

This idea motivated the development of a different key-finding algorithm that weights incoming tones according to the tonal hierarchy. Krumhansl and Schmuckler (Krumhansl 1990a) suggested this approach might result in a more accurate and efficient algorithm than simply assessing whether or not tones are scale members, as was done in the Longuet-Higgins and Steedman model. The input to the algorithm is the distribution of tones in the input segment weighted according to their duration. That is, it is based on the summation of tone durations (total duration) of each of the chromatic scale tones in the segment. The algorithm correlates this input with the 24 major and minor Krumhansl and Kessler (1982) key profiles. The correlations give a measure of the strength of each possible key. The algorithm was compared with that of Longuet-Higgins and Steedman (1971), with favorable results, and Cohen's (1991) listeners' judgments of key for the Bach Preludes and Fugues in the *Well-Tempered Clavier*. To visualize the results, they were projected onto the geometric map of keys that resulted from correlating the tonal hierarchies of different keys (Krumhansl and Kessler 1982; Krumhansl 1990a).

This led to the development of a dynamic version of the model using a self-organizing neural net model (Krumhansl and Toiviainen 2001; Toiviainen and Krumhansl 2003). It used data collected using a *continuous probe-tone task* in which the probe tone was sounded continuously while the music was played. Listeners heard the entire piece of music (a Bach organ piece) and rated how well one probe tone fit with the music as it progressed in time. Then another probe tone was chosen and the procedure was repeated until all 12 chromatic scale tones were used. This produces a probe-tone profile at each point in time during the music. These results were projected onto the self-organizing neural net resulting from

training it on the probe-tone results of Krumhansl and Kessler (1982), which replicated the earlier geometric map of keys. The perceived tonality was represented as continuously changing patterns of color on the self-organizing net, with color representing the strength of the keys. This was compared to an alternative model using just the tone distributions in the music. The listeners' data showed influences of tonality over and above the emphasis given tones by frequency and duration in the music reflecting knowledge of the style beyond tone distributions.

A major contribution in key-finding models comes from the work of Temperley (1999; 2001; 2007; Temperley and Marvin 2008). Temperley (1999) suggested a number of modifications of the Krumhansl-Schmuckler algorithm. One modification, based on music-theoretic considerations, was to modify the weights of the tones. Another is that a penalty was imposed for changing key from one input segment to the next. Finally, a retrospective reevaluation of key was permitted. In addition to comparing the model with that of Longuet-Higgins and Steedman (1971), Temperley carried out the most extensive test to date using a music theory textbook's analysis of key in a large number of pieces. This was done on a measure-by-measure basis. The modified model performed well in these applications. Temperley (2001) describes additional theoretical concerns, model modifications, and tests of the model. It should be noted that more recently Temperley (2007) introduced an alternative approach to key-finding that is based on the probability of sequences of tones in different keys and assigns key according to these probabilities (using Bayes' theorem from statistics).

The models reviewed here utilize tonal hierarchies for finding the key of a musical selection, where the selection can range from a few notes to entire sections of music. The models to date indicate the utility of this modeling approach in these applications. Once the key is identified, then the functions of the tones in the key are determined, allowing subsequent processing, such as analyzing the music for the harmony (chords).

3.4.10 Tonal Hierarchies and Musical Tension

As one listens to music, there are points in time, such as at phrase endings, where the music seems relatively complete and the feeling of tension is low. At other points in time, the music generates strong expectations that the music must continue onto some resolution, with high levels of tension. The rising and falling of tension is fundamental to the experience of music and may contribute to the emotional response (Meyer 1956; Krumhansl 2002; Krumhansl and Agres 2008). The question addressed in this final section is how a model might account for these variations in tension over time.

Lerdahl's (1988, 1996, 2001, 2009) *pitch space model* makes quantitative predictions of the how the degree of tension varies over time. The music is first analyzed as chords in keys (the standard harmonic analysis, such as a G major chord in the key of C major). The music analyzed in this way, called *events*, is the input to the model. The model

consists of four components, all of which contribute to predicted tension. Of most interest here is the *basic pitch space*, a theorized version of empirical measures of tonal hierarchies. The basic pitch space, shown in Fig. 3.2a, consists of five levels: the chromatic scale at the lowest level, the diatonic scale on the next level, then the triad level, the tonic-dominant level, and finally the tonic level.

To make quantitative predictions of the amount of tension caused by each event in the music, it is necessary to find a way of computing the distances between events. The distance between events is the sum of three numbers. The first, and most relevant here, is the number of changes that are needed to transform the basic pitch space for the first event into the basic pitch space for the second event. Take as an example the distance between I chord in C major (C major) and the vi chord (the chord built on the sixth tone of the F major scale, d minor). The basic pitch space for the C major chord is shown in Fig. 3.2a and the basic pitch space for the vi chord in F major (3 minor) is shown in Fig. 3.2b. There is one change on the diatonic level: B is changed to Bb. On the levels above this, C was represented on all five levels and is now represented on two levels (a difference of 3); E, which was represented on three levels is now represented on two levels (a difference of 1); and G, which was represented on four levels is now represented on two levels (a difference of 2). This gives a total of 7 for the number of changes to transform one basic pitch space into the other. To these are added the distances between chords on the circle of fifths for chords (2, that is, C to G, G to d) and the distance between keys (1, that is, C major to F major), for a total tension value of 10.

a

(a) Octave (root) level:	X												
(b) Fifths level:	X										X		
(c) Triad level:	X					X				X			
(d) Diatonic level:	X	X			X	X			X	X			X
(e) Chromatic level:	X	X	X	X	X	X	X	X	X	X	X	X	X
	C	C#	D	D#	E	F	F#	G	G#	A	Bb	B	

b

(a) Octave (root) level:	X												
(b) Fifths level:	X										X		
(c) Triad level:	X					X				X			
(d) Diatonic level:	X			X	X			X	X			X	
(e) Chromatic level:	X	X	X	X	X	X	X	X	X	X	X	X	X
	C	C#	D	D#	E	F	F#	G	G#	A	Bb	B	

Fig. 3.2 (a) The basic pitch space for the tonic triad chord (I) in C major. (b) The basic pitch space for the chord built on the second scale tone (ii) of F major. Seven elements change between the two basic spaces (See Lerdahl 2001; Lerdahl and Krumhansl 2007)

The distances as computed above are then applied to the prolongational structure, a tree representation that shows the way in which events in the music are related to one another. This is an instance of what Bharucha (1984) called an *event hierarchy*, as defined in the preceding text. The tree representation shows links that occur between events. It is important to note that these are sometimes nonadjacent in the music as shown in Fig. 3.3. The tree shows a link between the V chord (dominant, the major chord built on the fifth scale degree) and the I chord (tonic, the major chord built on the first scale degree, or tonic) although the ii chord (the minor chord built on the second scale degree) intervenes. In a tree such as this, the tension value for an event is the sum of the distances between the events at all levels up to the root of the tree. In Fig. 3.3, the ii chord is subordinate to the V chord and this is subordinate to the I chord at the root of the tree. The total distance for the ii chord from the root of the tree (its tension) is its distance from the V chord, plus the distance of the V chord from the I chord, with these distances computed as just described. To this value is added two other numbers, the surface dissonance (of each event) and attraction (between successive events) as specified in the model. This is the predicted tension.

To date, five studies have found empirical support for the model's account of tension (Bigand et al. 1996; Krumhansl 1996; Cuddy and Smith 2000; Smith and Cuddy 2003; Lerdahl and Krumhansl 2007). In the empirical studies, participants were asked to judge tension (either continuously as the music is played or, for the shorter segments, stopped at each successive point) and these judgments were compared with the model's quantitative predictions. Lerdahl and Krumhansl (2007) is the most extensive test of the model, using five excerpts written in quite different styles. The styles included traditional tonal-harmonic music, highly chromatic diatonic music, and two excerpts that might be analyzed in other scales, the six-tone hexatonic scale and the eight-tone octatonic scale. The predictions of the model

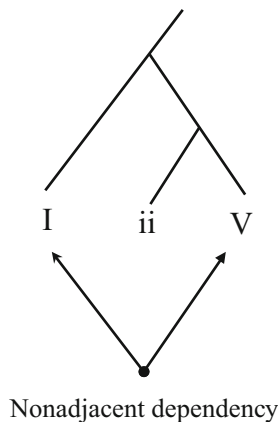


Fig. 3.3 Nonadjacent dependencies in Lerdahl's (2001) prolongation structure. The V chord links to the I chord even though the ii chord intervenes. The tension of the ii chord is computed as the distance between the ii chord and the V chord plus the distance between the V chord and the I chord

were strongly confirmed for all the excerpts. In some cases, modifications of the prolongational (tree) structure achieved a better fit between model and data, but the modifications were principled in terms of the pitch space model.

What is most relevant here is that the distance calculations in the Lerdahl (2001) model required no modification. Recall that these distances were calculated using the basic space that has essentially the same structure as the empirically measured tonal hierarchies. It is not obvious a priori that this would be the case. The experimental studies on tone, chord, and key distances on which the model is based tend to use short, schematic materials (e.g., scales, chords, chord cadences). These materials lack relationships found in extended musical excerpts between nonadjacent events as described in the prolongational structure. In some cases the nonadjacent events are quite distant in the music. Nonetheless, the distances calculated along the branches of the trees successfully predicted the tension judgments. Thus, it is notable that the pitch space distance, calculated from the basic pitch space, extends to complex segments of musical compositions. In addition, the judgment of tension in these experiments is quite different from the more direct judgments of structure used in the basic experimental studies as described earlier.

This is an example of a deeply theorized proposal stimulated in part by the empirical work on tonal hierarchy. It makes predictions that can be tested against data, which may lead to refinements of the theory. Other music theoretic models based on the empirical results on tonal hierarchy and related work may follow.

3.5 Tonal Hierarchy Theory and Implications for Cognitive Science

In this chapter, a theory of tonal hierarchies has been presented. It is a theory in the sense that it summarizes a large body of empirical evidence and makes predictions for empirical study. The theory asserts three propositions, one concerning the psychological status of the tonal hierarchy, one concerning the musical status of the hierarchy, and the third concerning the relationship between these subjective and objective descriptions of the tonal hierarchy.

The first proposition is that tonal hierarchies have psychological reality. Psychological interest in tonal hierarchies grew out of music-theoretic descriptions of Western tonal-harmonic music, the style that dominates Western music. In this style, certain tones are identified as more prominent, stable, and structurally significant than others. Thus, according to the theory, music establishes a hierarchy of tones. In Western music, the hierarchy is headed by the tonic. Other tones follow: those in the tonic triad, then the other scale tones, and finally the nonscale tones. A wide range of experimental methods have been developed to test whether the proposed hierarchies can be elicited in psychological experiments. The psychological reality of the tonal hierarchy is supported by converging evidence from numerous experimental studies of cognition, development, learning, neuropsychology, and cross-cultural psychology. Among its manifestations are tonal hierarchy's effects on

memory, the sense of stability and instability, the choice of phrase endings, the perceived relations between pitches, and the generation of melodic expectations.

Cross-cultural comparisons proved particularly important for sharpening the questions involved. For music from different cultures, listeners did not produce the Western tonal hierarchy but rather produced a tonal hierarchy consistent with the hierarchy of each musical culture. This general result argues against the view that tonal hierarchies can be explained by the harmonic structure (overtones) of complex tones. The argument acknowledges the fact that harmonic structure across cultures typically contains the perceptually privileged intervals of the octave and the perfect fifth. However, and significantly, if harmonic structure were the sole determinant of the tonal hierarchy then tonal hierarchies would not exhibit this kind of cross-cultural variability. Rather, these variations suggested that the tonal hierarchy is cognitive in origin and that it is internalized in large part through experience with music.

The second proposition is that tonal hierarchies are musical facts. As indicated earlier, the initial impulse for the empirical studies was the description in music theory of hierarchies of tones in tonal-harmonic music. However, this claim resides in a complex theory of tonal-harmonic music and technical vocabulary that would be unknown to nonmusicians. A musical property more readily perceptible is needed to explain the generality of the experimental results just summarized. The property that was identified was the relative emphasis given tones in the music, the assumption being that tones high in the tonal hierarchy would be emphasized on the surface of the music. An analytic method was needed for describing the relative salience of tones. It was quantified by determining the frequencies and/or durations of each tone in the music. A number of note-count studies have analyzed large corpora of Western tonal-harmonic music. The results vary somewhat, depending on the particular pieces in the corpus, but at a broad level the results correspond to the empirically measured tonal hierarchies. More narrowly focused studies consider particular styles of music or the way a composer writes music to make the tonal hierarchy evident. Analysis of styles other than tonal-harmony also shows surface emphasis by frequency and duration. This line of research has been aided by computer encoding of large numbers of pieces and tools to conduct such studies of tone distributions.

The evidence just described for the psychological reality of tonal hierarchies and their status as musical facts raises the fundamental question as to their relationship. This issue is articulated by the third proposition of the theory: statistically frequent patterns in music, and the consequent salience of certain tones, enable listeners to orient to tonal hierarchies. Moreover, the proposition predicts that listeners rapidly adapt to style-appropriate tonal hierarchies even if the style is unfamiliar.

That listeners use tone distributions to form a tonal hierarchy has now been shown by experiments with a wide range of styles. Probe tone judgments evoked in Western listeners by simple schematic key-defining contexts closely match the distribution of tones in tonal-harmonic music. Westerners exposed to music of other cultures, that of India and to some extent Bali, demonstrate that sensitivity to tone distributions in the contexts induces rapid assimilation of the relevant tonal hierarchies. Examination of

the music showed that tones high in the hierarchy tended to be repeated more often and for longer durations. In studies of nontonal Western music, sensitivity to tone distributions is also observed, but with some qualifications, especially in the case of 12-tone serial music. Finally, the correspondence between tone distributions and the perceived tonal hierarchy is also found when the distribution of tones is manipulated experimentally according to novel schemes. Some of the results suggest that musicians are somewhat more sensitive to the distributions of tones than nonmusicians. Finally, computational models provide additional supporting evidence. It has been shown that neural network models are able to abstract tonal hierarchies (as well as harmonic and key relations) from both schematic and more realistic musical inputs. In addition, processing of statistical properties of the musical surface (the frequencies of occurrence of tones and tone combinations) has provided the basis for successful modeling of tonality induction.

In sum, there is a strong basis in experimental results to support the first proposition of the theory, that tonal hierarchies have psychological reality. A metric of musical tonal hierarchies, the distribution of tone frequencies and/or duration, matches well the psychological tonal hierarchy. Thus, tonal hierarchies are musical facts, which is the second proposition of the theory. Finally, building on these two first propositions, the third proposition is that sensitivity to the distribution of tones in the music enables listeners to abstract the tonal hierarchy. The evidence supports this, and also that listeners can adapt quite rapidly to unfamiliar styles.

Viewed from a broader cognitive science perspective, the literature reviewed here demonstrates the operation of two general psychological principles. The first is the existence of a frame of reference to guide perception and cognition that takes the form of a hierarchy of pitches. Other perceptual and cognitive domains contain reference points; these reference points provide an economical description of the domain in question. In the context of music, the tones high in the hierarchy serve as reference points with respect to which other tones and chords are efficiently encoded and accurately remembered.

Despite the common principle of reference points, tonal hierarchies appear to be unique to music. Nothing analogous appears, for example, in language or in other perceptual domains. This raises the possibility that tonal hierarchies are especially important in music because most listeners (those without absolute pitch, the ability to name tones in isolation) process music relatively. In other words, musical tones do not have inherent qualities that are invariant across contexts. Instead, pitches are heard in context, and related to one another in that context. The tonal hierarchy provides a stable framework for establishing these relationships. Moreover, the representation of the hierarchy may become isolable at the neural level so that it may be selectively lost or spared in brain pathology.

The second general psychological principle that underlies tonal hierarchies is a mechanism for extracting environmental regularities through sensitivity to distributions of tones and tone combinations. This principle came into focus as the empirical question shifted from establishing the psychological reality of tonal hierarchies to investigating whether tonal hierarchies are learned and, if so, how. The evidence pointed to the involvement of sensitivity to tone distributions, specifically how the

tones are emphasized by frequency and duration. These measures could be readily quantified independently of the style. They were found to match well the empirical measures of tonal hierarchy. This surface emphasis appeared to be accessible to nonmusicians as seen by their judgments closely following the tone distributions. The results for listeners familiar with the style also reflected the tone distributions in addition to the influences of style-specific knowledge. This suggests that tone distributions and style-specific knowledge are complementary, rather than conflicting.

Music appears to be very distinct from other perceptual and cognitive domains. Moreover, a vast variety of different styles of music can be found across cultures and historical periods. This suggests it would be difficult to draw on other domains of psychological study or to make generalizations across musical styles. However, when viewed in terms of underlying cognitive principles, the commonalities with other domains are revealed at a deep level and the diversity of musical styles can be better understood.

References

- Aarden B (2003) Dynamic melodic expectancy. Unpublished doctoral dissertation, School of Music, Ohio State University.
- Auhagen W, Vos PG (2000) Experimental methods in tonality induction research: a review. *Music Percept* 17:417–436.
- Ayotte J, Peretz I, Hyde K (2002) Congenital amusia: a group study of adults afflicted with a music-specific disorder. *Brain* 125:238–251.
- 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.
- Besson M, Faïta F, Peretz I, Bonnel AM, Requin J (1998) Singing in the brain: independence of lyrics and tunes. *Psychol Sci* 9:494–498.
- Bharucha JJ (1984) Event hierarchies, tonal hierarchies and assimilation: a reply to Deutsch and Dowling. *J Exp Psychol Gen* 113:421–425.
- Bharucha JJ, Krumhansl CL (1983) The representation of harmonic structure in music: hierarchies of stability as a function of context. *Cognition* 13:63–102.
- Bigand E, Poulin-Charronnat B (2006) Are we “experienced listeners”? A review of the musical capacities that do not depend on formal musical training. *Cognition* 100:100–130.
- 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.
- Boltz M (1989a) Perceiving the end: effects of tonal relationships on melodic completions. *J Exp Psychol Hum Percept Perform* 15:749–761.
- Boltz M (1989b) Rhythm and “good endings”: effects of temporal structure on tonality judgments. *Percept Psychophys* 46: 9–17.
- Brown H, Butler B, Jones MR (1994) Musical and temporal influences on key discovery. *Music Percept* 11:371–407.
- Castellano MA, Bharucha JJ, Krumhansl CL (1984) Tonal hierarchies in the music of North India. *J Exp Psychol Gen* 113:394–412.
- Cohen AJ (1991) Tonality and perception: musical scales primed by excerpts from the Well-tempered Clavier of JS Bach. *Psychol Res/Psychol Forsch* 53:305–314.
- Cohen AJ (2000) Development of tonality induction: plasticity, exposure, and training. *Music Percept* 17:437–459.

- Cuddy LL (1997) Tonal relations. In: Deliège I, Sloboda J (eds), *Perception and Cognition of Music*. Hove, Sussex: Taylor & Francis, pp. 329–352.
- Cuddy LL (2000) Perception and representation of musical structure. Paper presented at the XXVII International Congress of Psychology, Stockholm, July 23–28.
- Cuddy LL, Badertscher, B (1987) Recovery of the tonal hierarchy: some comparisons across age and levels of musical experience. *Percept Psychophys* 41:609–620.
- Cuddy LL, Duffin JM (2005) Music, memory, and Alzheimer's disease: is music recognition spared in dementia and how can it be assessed? *Med Hypotheses* 64:229–235.
- Cuddy LL, Lunnery CA (1995) Expectancies generated by melodic intervals: perceptual judgments of melodic continuity. *Percept Psychophys* 57:451–462.
- Cuddy LL, Smith NA (2000) Perception of tonal pitch space and tonal tension. In: Greer D (ed), *Musicology and Sister Disciplines: Past Present Future*. Oxford: Oxford University Press, pp. 47–59.
- Cuddy LL, Cohen AJ, Mewhort DJK (1981) Perception of structure in short melodic sequences. *J Exp Psychol Hum Percept Perform* 7:869–883.
- Cuddy LL, Balkwill L-L, Peretz I, Holden RR (2005) Musical difficulties are rare: a study of “tone deafness” among university students. *Ann NY Acad Sci* 1060:311–324.
- Dalla Bella S, Peretz I (1999) Music agnosias: selective impairments of music recognition after brain damage. *J New Music Res* 28:209–216.
- DeVoto M (1986) Tonality. In: Randel DM (ed), *The New Harvard Dictionary of Music*. Cambridge, MA: Belknap, pp. 862–863.
- Dowling WJ (1978) Scale and contour: two components of a theory of memory for melodies. *Psychol Rev* 85:341–354.
- Dowling WJ (1999) The development of music perception and cognition. In: Deutsch D (ed), *The Psychology of Music* (2nd ed.) San Diego: Academic Press, pp. 603–626.
- Drayna D, Manichaikul A, de Lange M, Snieder H, Spector T (2001) Genetic correlates of musical pitch recognition in humans. *Science* 291:1969–1972.
- Fornazzari L, Castle T, Nadkarni S, Ambrose M, Miranda D, Apanasiewicz N, et al. (2006) Preservation of episodic musical memory in a pianist with Alzheimer disease. *Neurology* 66: 610–611.
- Frankland BW, Cohen AJ (1990) Expectancy profiles generated by major scales: group differences in ratings and reaction time. *Psychomusicology* 9:173–192.
- Garner WR, Hake HW, Eriksen CW (1956) Operationism and the concept of perception. *Psychol Rev* 63:149–159.
- Handel S (1989) *Listening: An Introduction to the Perception of Auditory Events*. Cambridge, MA: MIT Press.
- Hughes M (1977) A quantitative analysis. In: Yeston M (ed), *Readings in Schenker Analysis and Other Approaches*. New Haven: Yale University Press, pp. 144–164.
- Huron D (1993) Chordal-tone doubling and the enhancement of key perception. *Psychomusicology* 12:154–171.
- Huron D (2006) *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press.
- Huron D, Parncutt R (1993) An improved model of tonality perception incorporating pitch salience and echoic memory. *Psychomusicology* 12:154–171.
- Hyde K, Peretz I (2004) Brains that are out of tune but in time. *Psychol Sci* 15:356–360.
- Janata P, Reisberg D (1988) Response-time measures as a means of exploring tonal hierarchies. *Music Percept* 6:161–172.
- Järvinen T (1995) Tonal hierarchies in jazz improvisation. *Music Percept* 12:415–437.
- Jordan DS, Shepard RN (1987) Tonal schemas: evidence obtained by probing distorted musical scales. *Percept Psychophys* 41:489–504.
- Kessler EJ, Hansen C, Shepard RN (1984) Tonal schemata in the perception of music in Bali and in the West. *Music Percept* 2:131–165.
- Knopoff L, Hutchinson W (1983) Entropy as a measure of style: the influence of sample length. *J Music Theory* 27:75–97.

- Krumhansl CL (1979) The psychological representation of musical pitch in a tonal context. *Cogn Psychol* 11:346–374.
- Krumhansl CL (1985) Perceiving tonal structure in music. *Am Sci* 73: 371–378.
- Krumhansl CL (1987) Tonal and harmonic hierarchies. In: Sundberg J (ed), *Harmony and Tonality*. Stockholm: Royal Swedish Academy, pp. 13–32.
- Krumhansl CL (1990a) *Cognitive Foundations of Musical Pitch*. New York: Oxford University Press.
- Krumhansl CL (1990b) Tonal hierarchies and rare intervals in music cognition. *Music Percept* 7:309–324.
- Krumhansl CL (1991) Melodic structure: theoretical and empirical descriptions. In: Sundberg J (ed), *Music, Language, Speech and Brain*. London: Macmillan.
- Krumhansl CL (1995a) Effects of musical context on similarity and expectancy. *Systematische Musikwissenschaft (Systematic Musicology)* 3:211–250.
- Krumhansl CL (1995b) Music psychology and music theory: Problems and prospects. *Music Theory Spectrum* 17:53–80.
- Krumhansl CL (1996) A perceptual analysis of Mozart's piano sonata K. 282: segmentation tension and musical ideas. *Music Percept* 13:401–432.
- Krumhansl CL (2002) Music: a link between cognition and emotion. *Curr Dir Psychol Sci* 11:45–50.
- Krumhansl CL, Agres KA (2008) Musical expectancy: the influence of musical structure on emotional response. *Brain Behav Sci* 31:584–585.
- Krumhansl CL, Keil FC (1982) Acquisition of the hierarchy of tonal functions in music. *Mem Cogn* 10:243–251.
- Krumhansl CL, Kessler EJ (1982) Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychol Rev* 89:334–368.
- Krumhansl CL, Schmuckler MA (1986) The Petroushka chord: a perceptual investigation. *Music Percept* 4:153–184.
- Krumhansl CL, Shepard RN (1979) Quantification of the hierarchy of tonal functions within a diatonic context. *J Exp Psychol Hum Percept Perform* 5:579–594.
- Krumhansl CL, Toiviainen P (2001) Tonal cognition. *Ann NY Acad Sci* 930:77–91.
- Krumhansl CL, Sandell GJ, Sergeant DC (1987) The perception of tone hierarchies and mirror forms in twelve-tone serial music. *Music Percept* 5:31–78.
- Krumhansl CL, Louhivuori J, Toiviainen P, Järvinen T, Eerola T (1999) Melodic expectation in Finnish spiritual folk hymns: convergence of statistical behavioral and computational approaches. *Music Percept* 17:151–195.
- Krumhansl CL, Toivanen P, Eerola T, Toiviainen P, Järvinen T, Louhivuori J (2000) Cross-cultural music cognition: cognitive methodology applied to North Sami yoiks. *Cognition* 76:13–58.
- Lamont A, Cross I (1994) Children's cognitive representations of musical pitch. *Music Percept* 12:27–55.
- Lantz ME, Cuddy LL (1998) Total and relative duration as cues to surface structure in music. *Can Acoust* 26:56–57.
- Lantz ME, Kilgour A, Nicholson KG, Cuddy LL (2003) Judgments of musical emotion following right hemisphere damage. *Brain Cogn* 51:190–191.
- Leman M (2000) An auditory model of the role of short-term memory in probe-tone ratings. *Music Percept* 17:481–509.
- Lerdahl F (1988) Tonal pitch space. *Music Percept* 5:315–349.
- Lerdahl F (1996) Calculating tonal tension. *Music Percept* 13:319–363.
- Lerdahl F (2001) *Tonal Pitch Space*. New York: Oxford University Press.
- Lerdahl F (2009) Genesis and architecture of the GTTM project. *Music Percept* 26:187–194.
- Lerdahl F, Krumhansl CL (2007) Modeling tonal tension. *Music Percept* 24:329–366.
- Liégeois-Chauvel C, Peretz I, Babai M, Laguitton V, Chauvel P (1998) Contribution to different cortical areas in the temporal lobes to music processing. *Brain* 121:1853–1867.
- Longuet-Higgins HC, Steedman MJ (1971) On interpreting Bach. *Mach Intell* 6:221–241.
- Marin OSM, Pery DW (1999) Neurological aspects of music perception and performance. In: Deutsch D (ed), *The Psychology of Music* (2nd ed.) San Diego: Academic Press, pp. 653–724.
- Marmel F, Tillmann B (2009) Tonal priming beyond tonics. *Music Percept* 26:211–221.

- Meyer LB (1956) *Emotion and Meaning in Music*. Chicago: University of Chicago Press.
- Miyazaki K (1989) Absolute pitch identification: effects of timbre and pitch region. *Music Percept* 7:1–14.
- Narmour E (1990) *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model*. Chicago: University of Chicago Press.
- Oram N, Cuddy LL (1995) Responsiveness of Western adults to pitch-distributional information in melodic sequences. *Psychol Res* 57:103–118.
- Palmer C, Krumhansl CL (1987a) Independent temporal and pitch structures in perception 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.
- Parncutt R (1988) Revision of Terhardt's psychoacoustical model of the root(s) of a musical chord. *Music Percept* 6:65–94.
- Parncutt R (1989) *Harmony: A Psychoacoustical Approach*. Berlin: Springer-Verlag.
- Parncutt R (1994) Template-matching models of musical pitch and rhythm perception. *J New Music Res* 23:145–167.
- Patel AD (2008) *Music, Language, and the Brain*. Oxford: Oxford University Press.
- Peretz I (1993a) Auditory agnosia: a functional analysis. In: McAdams S, Bigand E (eds), *Thinking in Sound: The Cognitive Psychology of Human Audition*. Oxford: Oxford University Press, pp 199–230.
- Peretz I (1993b) Auditory atonalia for melodies. *Cogn Neuropsychol* 10:21–56.
- Peretz I (1996) Can we lose memory for music? A case of music agnosia in a nonmusician. *J Cogn Neurosci* 8:481–496.
- Peretz I (2008). Musical disorders: from behavior to genes. *Curr Dir Psychol Sci* 17:329–333.
- Peretz I, Kolinsky R, Tramo M, Labrecque R, Hublet C, Demeurisse G, Belleville S (1994) Functional dissociations following bilateral lesions of auditory cortex. *Brain* 117:1283–1301.
- Peretz I, Belleville S, Fontaine S (1997) Dissociations entre musique et langage après atteinte cérébrale: un nouveau cas d'amusie sans aphasie. *Can J Exp Psychol* 51:354–368.
- Peretz I, Gosselin N, Tillmann B, Cuddy LL, Gagnon B, Trimmer CG, Paquette S, Bouchard B (2008) On-line identification of congenital amusie. *Music Percept* 25:331–343.
- Perlman M, Krumhansl CL (1996) An experimental study of internalized interval standards of Javanese and Western musicians. *Music Percept* 14:95–116.
- Piston W (1987) *Harmony* (revised and expanded by M DeVoto). New York: WW Norton.
- Rosch E (1975) Cognitive reference points. *Cogn Psychol* 7:532–47.
- Rosch E (1978) Principles of categorization. In: Rosch E, Lloyd BB (eds), *Cognition and Categorization*. Hillsdale, NJ: Lawrence Erlbaum.
- 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.
- Rosch E, Mervis CB (1975) Family resemblances: studies in the internal structure of categories. *Cogn Psychol* 7:573–605.
- Rowe R (2000) Key induction in the context of interactive performance. *Music Percept* 17:511–530.
- Saffran JR, Griepentrog GJ (2001) Absolute pitch in infant auditory learning: evidence for developmental reorganization. *Dev Psychol* 37:74–85.
- Saffran JR, Aslin RN, Newport EL (1996a) Statistical learning by 8-month-old infants. *Science* 274:1926–1928.
- Saffran JR, Newport EL, Aslin RN (1996b) Word segmentation: the role of distributional cues. *J Mem Lang* 35:606–621.
- Saffran JR, Newport EL, Aslin RN, Tunick RA, Barrueco S (1997) Incidental language learning: listening (and learning) out of the corner of your ear. *Psychol Res* 8:101–105.
- Saffran JR, Johnson EK, Aslin RN, Newport EL (1999) Statistical learning of tone sequences by human infants and adults. *Cognition* 70:27–52.
- Schmuckler MA (1989) Expectation in music: investigation of melodic and harmonic processes. *Music Percept* 14:295–318.
- Shepard RN (1964) Circularity in judgments of relative pitch. *J Acoust Soc Am* 36:2346–2353.

- Shuter-Dyson R (1999) Musical ability. In: Deutsch D (ed), *The Psychology of Music* (2nd ed) San Diego: Academic Press, pp. 627–652.
- Smith NA, Cuddy LL (2003) Perceptions of musical dimensions in Beethoven's Waldstein Sonata: an application of tonal pitch space theory. *Musicae Scientiae* 7:7–34.
- Smith NA, Schmuckler MA (2004) The perception of tonal structure through the differentiation and organization of pitches. *J Exp Psychol Hum Percept Perform* 30:268–286.
- Speer JR, Meeks PU (1985) School children's perception of pitch in music. *Psychomusicology* 5:49–56.
- Steinke WR, Cuddy LL, Peretz I (1994) Dissociation of music and cognitive abstraction abilities in normal and neurologically impaired subjects. In: *Proceedings of the 3rd International Conference on Music Perception and Cognition*, Liège, Belgium pp. 425–426.
- Steinke WR, Cuddy LL, Holden RR (1997) Dissociation of musical tonality and pitch memory from nonmusical cognitive abilities. *Can J Exp Psychol* 51:316–334.
- Steinke WR, Cuddy LL, Jakobson LS (2001) Dissociations among functional subsystems governing melody recognition after right hemisphere damage. *Cogn Neuropsychol* 18:411–437.
- Temperley D (1999) What's key for key? The Krumhansl-Schmuckler key-finding algorithm reconsidered. *Music Percept* 17:65–100.
- Temperley D (2001) *The Cognition of Basic Musical Structures*. New York: Oxford University Press.
- Temperley D (2007) *Music and Probability*. Cambridge, MA: MIT Press.
- Temperley D, Marvin EW (2008) Pitch-class distribution and the identification of key. *Music Percept* 25:193–212.
- Terhardt E (1979) Calculating virtual pitch. *Hear Res* 1:155–182.
- Terhardt E, Stoll G, Seewann M (1982a) Pitch of complex signals according to virtual-pitch theory. Test examples and predictions. *J Acoust Soc Am* 71:671–678.
- Terhardt E, Stoll G, Seewann M (1982b) Algorithm for extraction of pitch and pitch salience from complex tonal signals. *J Acoust Soc Am* 71:679–388.
- Thompson WF (2008) *Music, Thought, and Feeling: Understanding the Psychology of Music*. New York: Oxford University Press.
- Thompson WF, Cuddy LL (1997) Music performance and the perception of key. *J Exp Psychol Hum Percept Perform* 23:116–135.
- Thompson WF, Mor S (1992) A perceptual investigation of polytonality. *Psychol Res* 54:60–71.
- Thompson WF, Cuddy LL, Plaus C (1997) Expectancies generated by melodic intervals: evaluation of principles of melodic implication in a melody completion task. *Percept Psychophys* 59:1069–1076.
- Tillmann B, Bharucha JJ, Bigand E (2000) Implicit learning of tonality: a self-organizing approach. *Psychol Rev* 107:885–913.
- Toivainen P, Krumhansl CL (2003) Measuring and modeling real-time responses to music: tonality induction. *Perception* 32:741–766.
- Trainor LJ, Trehub SE (1994) Key membership and implied harmony in Western tonal music: developmental perspectives. *Percept Psychophys* 56:125–132.
- Trehub SE (2000) Human processing predispositions and musical universals. In: Wallin NL, Merker B, Brown S (eds), *The Origins of Music*. Cambridge, MA: MIT Press.
- Trehub SE, Trainor LJ (1993) Listening strategies in infancy: the roots of music and language development. In: McAdams S, Bigand E (eds), *Thinking in Sound: The Cognitive Psychology of Human Audition*. Oxford: Oxford University Press, pp 278–327.
- Trehub SE, Schellenberg EG, Hill DS (1997) The origins of music perception and cognition: a developmental perspective. In: Deliège I, Sloboda J (eds), *Perception and Cognition of Music*. Hove, Sussex: Taylor & Francis, pp. 103–128.
- Vanstone AD, Cuddy LL (2010) Musical memory in Alzheimer disease. *Aging, Neuropsychol Cogn* 17:108–128.
- Vos PG, Leman M. (2000) Guest editorial: tonality induction. *Music Percept* 17:401–544.
- Whitfield IC, Evans EF (1965) Responses of auditory cortical neurons to stimuli of changing frequency. *J Physiol* 28:655–672.

- Yonelinas AP, Hockley WE, Murdock BB (1992) Tests of the list-strength effect in recognition memory. *J Exp Psychol Learn Mem Cogn* 18:345–355.
- Youngblood JE (1958) Style as information. *J Music Theory* 2:24–35.
- Zenatti A (1993) Children’s musical cognition and taste. In Tighe TJ, Dowling WJ (eds), *Psychology of Music: The Understanding of Melody and Rhythm*. Hillsdale, NJ: Lawrence Erlbaum, pp. 177–196.

Chapter 4

Music Acquisition and Effects of Musical Experience

Laurel J. Trainor and Kathleen A. Corrigan

4.1 Introduction

Rather little is known about how children acquire musical knowledge. However, everyday exposure to the music of one's culture does lead to implicit knowledge about its pitch and rhythmic structure, just as exposure to a particular language leads to implicit knowledge about its structure. While all children attend school with the goal of becoming literate, some children engage in formal music training whereas others do not. Thus music offers the opportunity to compare the effects of a wide range of experiences (Trehub and Trainor 1998).

This chapter examines the effects of musical experience on the development of three aspects of musical structure: pitch organization, rhythm, and emotional expression. For each aspect of musical structure, this chapter (1) selectively overviews what is known about how it develops, in cross-cultural perspective where possible; (2) examines how adult musicians and nonmusicians differ (the end-state of development); and (3) reviews the evidence for the role of experience in causing these differences. Finally, the effects of musical experience on other cognitive domains are considered in relation to whether it has general benefits across many domains or specific benefits in a few domains such as reading or visual-spatial skills, and what the mechanisms might be for such transfer.

Engaging infants and young children in music appears to be a cross-cultural universal, and it has been suggested that music, like language, is a species-specific behavior important for complex human social interaction (e.g., Trehub 2000, 2003). As recorded music becomes easier to produce and distribute, the nature of human musical engagement is changing, and the speed at which new musical compositions and styles are evolving appears to be increasing. In this context, it is important to consider the nature of musical development and the effects of different kinds of musical experience.

L.J. Trainor (✉)

Department of Psychology, Neuroscience & Behaviour, McMaster University,
Hamilton, ON, Canada, L8S 4K1
e-mail: LJT@mcmaster.ca

4.2 Pitch Organization

4.2.1 *Development of Pitch Organization: Musical Enculturation*

Musical pitch structure includes several different aspects that, although they interact, have different developmental trajectories. The most basic level is that of *individual musical tones*, which typically have energy at a fundamental frequency and integer multiples of that frequency, called harmonics (e.g., for a fundamental frequency of 100 Hz, there would also be energy at 200, 300, 400, ... Hz). The auditory system analyzes this spectrotemporal information and extracts a pitch percept that normally corresponds to the fundamental frequency. Individual tones are concatenated into sequences to form *melodic structures* and they are also combined simultaneously to form chords which are put into sequences to form *harmonic structures*. The details and indeed the presence of melodic and harmonic structure vary substantially between musical systems, but some basic principles appear to be close to universal, such as (1) to give prominence to pitch intervals (the pitch distance between tones) whose fundamental frequencies stand in small integer ratios and sound consonant; (2) to treat tones an octave apart as functionally equivalent (e.g., all the Cs or all the Ds on the piano); (3) to divide the octave into a small number of notes (usually between five and nine) that serve as discrete pitches for musical composition; (4) to have two or more interval sizes in scales so that the notes of the scale can be differentiated and take on different functions; and (5) to favor relative over absolute pitch representations, allowing melodies to be recognized across different starting pitches.

Because so little is known about children's acquisition of non-Western musical systems, this section on pitch organization focuses on the acquisition of Western tonal music. It will be for future research to determine whether the principles by which Western musical knowledge is acquired apply universally. The first subsection examines basic pitch processing and enculturation to Western melodic and harmonic structure under conditions of normal everyday exposure. Later subsections examine the effects of specific musical experience and training on musical acquisition.

Very young infants have some ability to distinguish pure tones with different frequencies. For example, during the last month before birth, the fetus responds to a change in pitch of roughly an octave (Hepper and Shahidullah 1994; Lecanuet et al. 2000) and neonates show event-related potential (ERP) responses in electroencephalograph (EEG) recordings to a 10% change in the pitch of a 1,000-Hz pure tone (Leppänen et al. 1997, 2004; Čeponienė et al. 2002). Pitch discrimination improves rapidly over the months after birth, although it does not reach adult levels until 8–10 years of age (Werner and Marean 1996). However, the ability to discriminate the smallest meaningful pitch difference in Western music, the semitone (6%), is present at least as early as 2 months of age (Werner and Marean 1996; He et al. 2009).

Musical tones are typically complex, containing harmonics that the auditory system integrates into a percept of a single tone with a particular timbre. The ability

to extract pitch from complex tones can be measured by the ability to perceive the pitch of the missing fundamental. If all energy is removed at the fundamental frequency, the pitch of a complex tone does not change (although the timbre does) because the pitch is implied by the spectrotemporal pattern across the harmonics; hence, the phenomenon is referred to as perceiving the pitch of the missing fundamental. Behavioral data indicate that 7-month-old infants integrate the harmonics of complex tones into a single pitch percept (Clarkson and Clifton 1995), and recent EEG data indicate that cortical representations for the pitch of the missing fundamental are present at 4 months of age (He and Trainor 2009). Thus, young infants have the basic pitch perception capabilities to allow musical processing.

Indeed, 2-month-old infants are able to recognize familiar melodies (Plantinga and Trainor 2009) and even neonates can segregate tones in a sequence into those with high pitch and those with low pitch (Winkler et al. 2003). Further, infants are sensitive to a number of melodic structural features that are similar across musical systems. First, they more readily encode scales with unequal interval sizes than scales with only one size of interval (Trehub et al. 1999). Second, infants find it easier to process certain musical intervals than others. Specifically, like adults, they are better able to process consonant than dissonant intervals. Intervals that sound consonant (pleasant to adults) consist of tones for which fundamental frequencies stand in small integer ratios, such as 2:1 (octave) and 3:2 (perfect fifth) whereas intervals with larger integer ratios, such as 15:8 (major seventh) or 45:32 (tritone) sound dissonant (rough or unpleasant). Given a set of consonant-interval standards, infants readily detect occasional dissonant comparison intervals, but they are unable to do the reverse discrimination (i.e., detect occasional consonant intervals among a set of dissonant intervals; Trainor 1997). Similarly, Schellenberg and Trainor (1996) found that infants more readily detect a change in a perfect fifth that creates a dissonant interval than a change that creates a different consonant interval, indicating that consonant intervals sound similar to infants. Further, infants as young as 2 months of age prefer to listen to consonant intervals compared to dissonant intervals (Trainor et al. 2002a). The superior processing of consonant intervals extends to melodic processing. Infants are better able to detect changes to melodies with prominent consonant intervals compared to melodies with prominent dissonant intervals (Trainor and Trehub 1993).

A third melodic structural feature to which infants are sensitive is transpositional invariance, reflecting the fact that melodies maintain their identity when played at higher or lower pitch levels as long as the relative pitch distances between tones remain constant. Although infants may process absolute pitch under some circumstances (Saffran and Griepentrog 2001; Volkova et al. 2006), they appear to favor relative pitch representations, as do most adults (Trehub et al. 1984; Plantinga and Trainor 2005, 2008). One indication that absolute pitch information for isolated tones fades rapidly in adults who do not possess absolute pitch is that the more interference or distractor tones with random pitches that are placed between two target tones, the worse adults are at detecting whether or not the target tones have the same or different pitches (Ross et al. 2004). Plantinga and Trainor (2008) showed that the same is true for infants. On the other hand, when a melody is transposed

up or down in pitch, the absolute pitch of every note in the melody changes. The ability to recognize a melody in transposition therefore relies on processing the distances between tones of the melody, that is, relative pitch. Infants can detect a change to one note of a melody, even when the comparison melodies are presented in transposition to different pitch levels (e.g., Trehub et al. 1984; Trainor and Trehub 1992a), suggesting good relative pitch processing early in development. Long-term memory representations also appear to be primarily in terms of relative pitch in infants. After being exposed to a melody for a week, infants prefer to listen to an unfamiliar melody, and this preference is unaffected by whether the familiarized melody is presented at the pitch level heard during familiarization or at a new and substantially different pitch level (up or down a perfect fifth or tritone), indicating that long-term memory representations are coded in terms of relative pitch (Plantinga and Trainor 2005).

Despite these precocious abilities, however, it takes some time for young infants to become enculturated to Western major scale structure, despite daily exposure during everyday activities. Lynch et al. (1990) showed that Western infants were equally able to detect changes to unfamiliar Balinese scales and Western scales, although their parents performed much better with the Western scales. Trainor and Trehub (1992a) showed that although Western adults find it much easier to detect a change to an unfamiliar melody when the changed note deviates from notes of the major scale (or key) compared to a change that remains within the notes of the scale, infants are able to detect both types of changes, actually outperforming adults in some conditions involving within-key changes. Out-of-key notes are readily detected by adults because their implicit knowledge of Western scale structure makes such notes sound wrong, whereas within-key notes do not violate Western scale structure. Thus infants' relatively good performance on within-key changes actually indicates a lack of knowledge about Western scale structure. The average age at which scale knowledge, or key membership, is solidified remains unknown, but it is certainly present by 4 or 5 years of age (Trehub et al. 1986; Trainor 2005; Corrigan and Trainor 2009).

Cortical representations for melodies remain immature for a much longer period than representations for individual tones. The simultaneous activation of groups of neurons in auditory cortex in response to the presentation of a sound can be measured at the surface of the scalp through the electrical fields that are generated. Following the direction of axons between layers in auditory areas located around the Sylvian fissure, such neural events appear as dipolar patterns on the scalp, with anterior negativities concurrent with posterior positivities, or vice versa. The stages of sound processing can be tracked through a series of frontally positive and negative components in the ERP over time (e.g., see Luck 2005). During the first months after birth, ERP responses to sound are dominated by slow waves that are not present in adult responses (for a review, see Trainor 2007). In adults, cortical memory traces for sound can be examined by measuring responses to occasional changes (deviants) in a repeating sound (standard), or changes in the category of a stream of sounds. Such occasional deviants elicit a negative ERP component not present in the response to standards, termed a mismatch negativity (MMN; for reviews,

see Picton et al. 2000; Näätänen et al. 2007). The MMN typically peaks between 130 and 250 ms after the onset of the change depending on stimulus complexity, and appears at the scalp as a negativity at anterior sites, concurrently with a positivity at posterior sites, consistent with a primary generator of the electrical field in secondary auditory cortex. Cortical memory traces for infants are somewhat different. During the first couple of months after birth, no MMN is seen, but a simple change in the pitch of a repeating musical tone elicits an increase in a frontally positive slow wave. The amplitude of this wave decreases with age, and an MMN resembling that of adults emerges around 3 months of age (He et al. 2007, 2009). Adults also show MMN responses to more complex stimuli such as melodies, even when those melodies are presented in transposition from repetition to repetition (e.g., Fujioka et al. 2004). However, in infants as old as 6 months, changes to melodies in transposition produce an increase in a slow frontally positive wave rather than the adult frontally negative responses (Tew et al. 2009). In sum, despite the fact that young infants recognize melodies, cortical processing remains immature and it takes considerable musical exposure to become acculturated to the musical scales in the environment.

Elaborate harmonic structure is relatively rare across the world's musical pitch systems, and from this perspective, it is interesting that sensitivity to harmony is typically rather late in developing, not reaching adult levels until at least 12 years of age (Costa-Giomi 2003). However, the perceptual distinction between consonance and dissonance is likely a necessary precursor of sensitivity to harmony because chords are built primarily on consonant relations between the notes comprising them. For example, the two most prominent chords in a key, the tonic chord, which is based on the first note of the scale, and the dominant chord, which is based on the fifth note of the scale, involve simultaneous tones that form consonant intervals (see Tramo et al. 2001). Infants prefer to listen to consonant compared to dissonant intervals (Trainor and Heinmiller 1998; Zentner and Kagan 1998; Trainor et al. 2002a). Further, this preference for consonance may be innate as it is present even in hearing newborns of deaf parents (Masataka 2006). Using EEG measures, Koelsch et al. (2003) showed that when the final chord in a sequence (i.e., in a chord progression) contains a note that deviates from the key of that sequence, a brain response to this unexpected chord is elicited from children as young as 5 years of age. Similarly, children as young as 6 years were found to be faster to make judgments about the last chord in a progression of chords (e.g., they judged which of two vowels was sung on the last chord) when the final chord was a tonic chord compared to when it was a subdominant chord (based on the fourth note of the scale; Schellenberg et al. 2005). Although all of the notes of the tonic and subdominant chords are contained within the key, sequences ending in a subdominant chord sound incomplete to a Western-enculturated listener. In a recent study, Corrigall and Trainor (2009) found that children as young as 4 years rated sequences ending in a tonic chord as sounding "good" significantly more often than sequences ending in a subdominant chord. Thus some sensitivity to harmonic progressions is present as young as 4 years of age.

In Western music, melodic and harmonic aspects of pitch structure interact, such that a melody presented alone “implies” the harmony that could accompany it. Trainor and Trehub (1994) showed that adults and 7-year-olds, but not 5-year-olds, process melodies according to their implied harmony. Specifically, a change in one note of a melody that remained within the key of the melody, but implied a different harmony from the original melody, was readily detected by 7-year-olds and adults, but not by 5-year-olds. On the other hand, all three groups were sensitive to scale structure (keys) as they all readily detected changes that deviated from the established key. These results are consistent with those from probe tone studies wherein children rate how well a tone fits into a preceding context (e.g., Krumhansl and Keil 1982; Speer and Meeks 1985; Cuddy and Badertscher 1987).

In sum, young infants are able to derive pitch from harmonics, discriminate musical pitches, and differentiate consonant and dissonant intervals. On the other hand, it takes considerable musical exposure and maturation before sensitivity to culture-specific scale and harmonic structure emerges. It is noteworthy that none of these developments requires formal musical training or explicit knowledge of musical structure, but rather they arise through everyday exposure to music.

4.2.2 Differences Between Adult Musicians and Nonmusicians in Musical Pitch Processing

By the time they reach adulthood, musicians have spent many years practicing their instrument, often for several hours each day, and this experience typically began at an early age. Further, although music is based in the auditory modality, learning a musical instrument provides intense multisensory experience that integrates motor processing with auditory, visual, tactile, and proprioceptive input. To examine the effects of this experience, a number of studies have compared auditory and multisensory processing in adult musicians and nonmusicians to study the effects of experience. Although innate factors influencing the decision to engage in musical training at a young age often cannot be ruled out, these studies provide a good starting point for examining the effects of musical experience.

Magnetic resonance imaging (MRI) studies indicate structural brain differences between musicians and nonmusicians (see Bermudez et al. 2008; Schlaug 2009). For example, Schneider et al. (2002) found that gray matter in auditory cortex is enlarged in musicians compared to nonmusicians, and that the degree of enlargement is correlated with musical skill. Structural differences are not limited to the auditory cortex. Compared to nonmusicians, musicians have more gray matter in Broca’s area (Sluming et al. 2002), cerebellum (Hutchinson et al. 2003), and motor areas (Gaser and Schlaug 2003; Bangert and Schlaug 2006). Using a new measure of cortical thickness, Bermudez et al. (2008) found that the cortex of musicians is thicker than that of nonmusicians in secondary auditory cortical areas, particularly on the right, as well as in dorsolateral frontal cortex, an area associated with executive functions and working memory. Functional magnetic resonance imaging

(fMRI) studies also suggest the involvement of a wide network of areas (e.g., Koelsch and Siebel 2005). The involvement of frontal areas likely reflects the great demands that musical performance places on the retention, monitoring, and retrieval of sound patterns.

While MRI studies give detailed information about where processing takes place in the brain, the stages of sound processing can be tracked with EEG and magnetoencephalographic (MEG) recordings, which monitor changes in electrical and magnetic fields associated with the synchronous depolarization and firing of groups of neurons. The presentation of a sound triggers a series of positive and negative field deflections at the surface of the head, each reflecting activity at a particular time from one or more brain regions. Many ERP components have been found to be larger in amplitude and/or earlier in musicians compared to nonmusicians (see Näätänen et al. 2007; Trainor and Zatorre 2009 for reviews). For example, preattentive middle-latency responses originating in primary auditory cortex are enhanced in musicians (Schneider et al. 2002; Shahin et al. 2004). Similarly, several preattentive components originating in secondary auditory cortical areas are larger and/or earlier in musicians than in nonmusicians (e.g., N1b at 100 ms after stimulus onset, Pantev et al. 1998; N1c around 170 ms, P2 around 200 ms, Shahin et al. 2005; Kuriki et al. 2006). These differences reflect either an increased number of neurons involved in processing musical sounds or an increased synchronization across neurons in musicians compared to nonmusicians.

Another preattentive ERP component, the MMN, is also larger in musicians than in nonmusicians. MMN is activated when an unexpected sound is inserted into a stream of similar sounds. It is thought to reflect the updating of sensory memory (e.g., Picton et al. 2000; Näätänen et al. 2007). A pitch change in a short melody that is repeated in transposition from trial to trial elicits an MMN that is larger in musicians than in nonmusicians (Fujioka et al. 2004). Likewise, when two melodies are presented at the same time in a polyphonic texture, separate memory traces are formed for each melody, and the MMN elicited by changes in each melody is larger in musicians than in nonmusicians (Fujioka et al. 2005). For harmonic processing, Koelsch and his colleagues have shown that chords that are unexpected in the context elicit an early right anterior negativity that again is larger in musicians than in nonmusicians (Koelsch et al. 2002). Finally, ERP components that involve attentional processing are also larger in musicians compared to nonmusicians, as evidenced by a larger P3a component, reflecting attentional capture of sounds in an unattended stream (e.g., Fujioka et al. 2004, 2005); a larger P3b component, reflecting conscious decision making about a sound (e.g., Trainor et al. 1999); and larger gamma band responses, reflecting networks for focused attention (Shahin et al. 2008).

The previous two paragraphs outline substantial evidence for structural and functional differences between musicians and nonmusicians. But just how different are they? Do musicians and nonmusicians differ simply in the degree of cortical network activation or do they exhibit qualitatively different processing? The answer depends on the perspective taken. Although brain regions associated with musical processing are larger in musicians, and although ERP responses to musical sounds and violations of musical structure are generally larger and earlier in musicians, the

particular brain areas involved and the particular ERP components generated are the same in the two groups. Thus, one interpretation is that all people (in the absence of congenital or acquired amusia) are musical, and training simply enhances musical processes. Indeed, this was the conclusion of Trainor et al. (2002b) when they found preattentive brain responses in nonmusicians in response to out-of-key notes in a melody, even though the melody was transposed to a different key from repetition to repetition. Interestingly, some behavioral studies have led to similar conclusions. Bigand, Tillmann, and colleagues have shown that when implicit tasks are used so that participants do not have to make a musical judgment directly, nonmusicians are sensitive to key structure and harmonic structure (Bigand and Poulin-Charronnat 2006). For example, when asked to judge the tuning (e.g., Bharucha and Stoeckig 1986, 1987), consonance (Bigand and Pineau 1997; Tillmann et al. 1998; Bigand et al. 1999) or timbre (Tillmann et al. 2006) of the last chord in a sequence, nonmusicians as well as musicians are faster when this final chord is expected, given the preceding context, than when it is not expected.

Perhaps musician/nonmusician differences can best be viewed as follows. Through everyday listening experience, nonmusicians and musicians are exposed to similar culture-specific musical pitch structures, and both groups become sensitive to these structures, likely through the operation of automatic statistical learning mechanisms. On the other hand, with formal training, compared to nonmusicians, musicians amass more experience with music than nonmusicians, they learn complex motor-auditory interactions in order to play their instruments, and they typically acquire explicit as well as implicit knowledge about musical structure. Thus they build larger and faster networks than nonmusicians for processing musical structure.

The final question to consider in this section is whether the differences seen in musical processing between musician and nonmusician adults reflect experience or whether those who became musicians engaged in extensive musical training from a young age because they were genetically predisposed to process music easily. This question is very difficult to answer. However, a couple of lines of evidence suggest that musician/nonmusician differences in adulthood are mediated at least to some extent by experience. First, within musician groups, the size of the N1 response correlates negatively with the age of onset of music lessons (Pantev et al. 1998), as does the size of the P3 response (Trainor et al. 1999). Second, ERP enhancements are most pronounced for tones of the timbre of the musical instrument of practice. For example, pianists show larger N1 responses to piano tones than to trumpet tones, and trumpet players show the reverse (Pantev et al. 2001). Third, many of the ERP components that are larger in musicians compared to nonmusician adults remain somewhat neuroplastic and can be modified in amplitude or latency through laboratory training in adults. For example, Bosnyak et al. (2004) trained adult nonmusicians in frequency discrimination and found that behavioral improvements in discrimination were accompanied by increases in P2 amplitude that were specific to the trained frequency. Further, multisensory training can effect larger changes in auditory areas than auditory

training alone. Lappe et al. (2008) trained one group of nonmusicians to play simple note sequences on the piano. A second group heard the same sequences and made judgments about them without learning to play them. After this training, the group that experienced the multisensory training showed larger MMN responses to wrong notes in similar sequences.

In sum, the evidence suggests that adult musician/nonmusician differences reflect, to a considerable extent, different musical experiences in childhood. In the next section, we consider studies with children that have tested the role of experience more directly.

4.2.3 Effects of Formal Musical Training on Children's Perception of Pitch Structure

Surprisingly, little scientific research has examined the effects of musical training on musical development in infants and young children. Much of the existing literature concerns absolute pitch training, which is generally not considered a core musical ability. Here, we focus on the effects of formal musical training on enculturation to musical pitch structure, with an emphasis on the development of harmonic sensitivity.

A few studies have shown differences in brain responses in young children engaging in music lessons compared to children not taking lessons. Interestingly, regardless of musical training, the auditory cortex has a very long developmental trajectory. ERP responses from auditory cortex to isolated musical tones continue to mature well into the teenage years (Ponton et al. 2000; Trainor et al. 2003; Shahin et al. 2004). Specifically, ERP components at around 50 ms (P1), 100 ms (N1), and 200 ms (P2) after sound onset increase in amplitude and decrease in latency until about 10 years of age, and then decrease in amplitude until adult levels are reached at around 18 years of age. Early musical training affects this trajectory. Shahin et al. (2004) found that 4- and 5-year-old children taking music lessons showed ERP responses that were similar to those of children 2–3 years older who were not taking music lessons. Further, the responses were consistent with the effects of musical training being specific to the timbre of the musical tones of the instrument of practice. Shahin et al. (2008) analyzed the ERP data of Shahin et al. (2004) in the frequency domain, specifically looking at responses in the gamma band range (40–100 Hz). Induced or non-phased-locked gamma band responses are particularly interesting because they have been linked to top-down processing or executive functions relating to attention and memory. The results showed that induced gamma band responses were present only in the group engaging in music lessons, and in that group, only after a year of music lessons. These data converge with those of Fujioka et al. (2006), who used MEG to show that in children of this age, an ERP component (the N2), which is related to auditory attention and memory processes, matures differently over the course of a year in children taking music

lessons compared to children not taking music lessons. In sum, music lessons appear to affect basic auditory processing of isolated musical tones.

One study has examined the effects of musical training on brain responses to violations of Western harmonic structure. Jentschke et al. (2005) compared 11-year-old children in the Saint Thomas Boys Choir in Leipzig with children not engaging in formal musical training who were matched for IQ and parents' education level. They measured ERP responses to the final chord in a sequence, specifically examining an early right anterior negative component (ERAN) which is known to occur in response to musically unexpected chords (Koelsch et al. 2000), and found that the ERAN was larger in the musically trained group than in the untrained group. Thus, some of the brain differences seen in harmonic processing between adult musicians and nonmusicians are present at least as early as 11 years of age.

Very little is known about the effects of musical training on enculturation to scales and harmony in preschool children. To address this question, Corrigan and Trainor (2009) tested two groups of 4- to 5-year-old children, the first of which had no formal music training and the second of which was just beginning music lessons at the time of the initial test. At the second test, about 1 year later, the first group still had no musical training, but the children in the second group had studied an instrument for 1 year. Of most interest in the present context, key membership and harmony perception were studied by presenting a sequence of five chords that ended (1) on the tonic chord as expected by the rules of Western harmony (standard), (2) on an out-of-key chord (tonic minor), or (3) on a chord that was within the key but not in the expected harmony at that point (subdominant instead of tonic). Children judged whether each sequence (standard, out-of-key, out-of-harmony) was a "good" or "bad" rendition from a puppet sitting in front of them. At the first test, all children rated the out-of-key chords and out-of-harmony endings more often as "bad" compared to the standard ending, providing evidence that children as young as four have some knowledge of key membership and harmony. However, of most interest, at the second measurement 1 year later, the group taking music lessons performed significantly better than the group not taking lessons. Thus, musical training in the preschool period leads to faster acquisition of harmonic sensitivity.

In sum, both behavioral and brain-based measures indicate that children who take formal music lessons develop sensitivity to culture-specific musical features such as scales and harmony at an earlier age than children not engaged in musical training.

4.2.4 Summary of the Development of Musical Pitch Acquisition

Certain universal musical features, such as the harmonic structure of pitch and consonance and dissonance are processed by very young infants. However, it takes several years for children to acquire system-specific knowledge of scale structure and harmony. Formal musical training is not necessary for this as people acquire

system-specific knowledge of musical pitch structure through passive everyday exposure. At the same time, the research indicates that those with formal musical training in childhood develop enhanced processing for musical pitch as reflected in superior perceptual discrimination as well as in enhanced brain structures for processing music and functional brain responses to musical pitch. There is some suggestion that there might be a sensitive period for musical pitch acquisition that ends around 10–12 years, but this evidence is far from conclusive. However, it has been established that robust effects of musical training can be seen already in pre-school children who engage in learning to play a musical instrument.

4.3 Rhythm

4.3.1 *Development of Metrical Perception: Enculturation*

It could be argued that rhythm is the most fundamental aspect of music – there are many styles of music with little or no pitch variation or structure, but few musical styles without a temporal organization. Rhythms consist of sequences of sound events and silences. A prominent hypothesis is that the brain uses two basic perceptual organizational processes to encode, remember, retrieve, and produce rhythmic patterns (e.g., Lerdahl and Jackendoff 1983). One is *grouping*, whereby the beginnings and ends of phrases and subphrases are determined. The second, which will be the focus in the present chapter, is the derivation of *metrical* structure. Listeners use the context of onset intervals between successive sound events, in conjunction with the duration, intensity, and pitch of sound events, to extract an ongoing metrical beat hierarchy (Jones and Boltz 1989; Large and Jones 1999). For Western music, beats at each level of the hierarchy are typically evenly spaced in time, and higher levels of the hierarchy are formed by combining every two or every three beats of the previous level, and lower levels of the hierarchy are formed by dividing each beat of the previous level into two or three beats. Metrical structure is not given directly in the stimulus, but is derived in the brain. Indeed, beats can be perceived when there is no sound event at all, and beats can be derived even when the loudest and longest sound events are off the beat, as in syncopation. At the same time, metrical extraction follows orderly rules, and even musically untrained adults show considerable agreement as to where the beats are in music, as indicated by their tapping behavior (e.g., Drake et al. 2000a; Snyder and Krumhansl 2001; Repp 2005). Adults appear to use statistical regularities in the input to extract the meter (e.g., Hannon et al. 2004).

The fundamental importance of metrical extraction for musical behavior is evident in the fact that this ability is what allows people to sing, dance, and play musical instruments together in synchrony. It also distinguishes humans from most other species. Recent evidence suggests that the few species that are able to synchronize to an external auditory beat are those who are also capable of vocal imitation (Schachner

et al. 2008). This suggests that metrical structure is likely fundamental to the complex communication systems – music and language – that have evolved in humans.

In some respects, infants are precocious rhythm processors. As young as 2–5 months of age, infants are able to discriminate simple rhythm patterns (Chang and Trehub 1977; Demany et al. 1977). By 2 months, infants can discriminate the tempi of isochronous beat patterns, and show optimal discrimination around 600 ms onset-to-onset (Baruch and Drake 1997), which is similar to adults. As young as 7–9 months, infants are able to recognize rhythms across variations in tempo and frequency (Trehub and Thorpe 1989). At 6 months, infants use duration cues for grouping successive sound events into phrases (Trainor and Adams 2000). Finally, a recent EEG study revealed that even newborn infants can extract a regular beat structure from temporal patterns (Winkler et al. 2009). Specifically, in the context of a rhythmic pattern, omission of an expected downbeat produced an ERP component in the newborns that is associated with violation of expectation.

Infants can extract metrical structures at 6 months of age (Morrongiello 1984; Hannon and Trehub 2005a). They can use statistical properties to categorize rhythm patterns where sound events are more likely to occur on every second beat from patterns where sound events are more likely to occur on every third beat (Hannon and Johnson 2005). Further, at 9 months infants detect changes in pitch or timing more readily in sequences with strong metrical structures than in sequences with weak metrical structures, indicating that metrical structure aids in encoding and processing rhythmic patterns in infancy (Bergeson and Trehub 2006).

The experience of metrical structure is intimately tied to the experience of rhythmic movement. Adults often feel a desire to move with the beat when listening to rhythmic music. There is likely a genetic basis for the interaction of movement and auditory rhythms, but there is also evidence for a learned aspect in that a person's preferred auditory beat tempo is related to his or her speed of walking (Todd et al. 2007). It has been suggested that this strong connection between auditory and movement rhythms might arise ontogenetically as the fetus and the young infant experience correlated sound and movement as they are walked, bounced and rocked (Hannon and Trainor 2007; Trainor 2007, 2008). Phylogenetically, rhythmic movement arose long before hearing, and can be seen in species as evolutionarily ancient as jellyfish.

Phillips-Silver and Trainor (2005) have argued that not only does music make people want to move, but movement can also influence how people experience a metrical structure. They presented 7-month-old infants with an ambiguous rhythm pattern that could be interpreted as being in either duple or in triple meter. Specifically, the rhythm pattern was six beats long and contained no accents other than the beginning of each six-beat group. Thus, it was metrically ambiguous as to whether the six-beat pattern was composed of three groups, each containing two beats (as in a march) or two groups, each containing three beats (as in a waltz). While listening to the ambiguous pattern, one group of infants was bounced in the arms of an experimenter on every second beat whereas another group was bounced on every third beat. After this experience linking movement to the ambiguous rhythm, those infants who were bounced on every second beat preferred to listen to a version of the rhythm with intensity accents added every second beat (as in a march) whereas those infants who

were bounced on every third beat preferred to listen to a version with accents added on every third beat (as in a waltz). Because infants all heard the same ambiguous rhythm during familiarization, the preference differences indicate that infants in the two groups perceived the ambiguous rhythm as being in different metrical structures. Thus, bouncing on every second beat caused them to perceive it as a march and bouncing on every third beat caused them to perceive it as a waltz. Similar results were found for adults (Phillips-Silver and Trainor 2007).

Young infants are motorically immature, and they were bounced in the arms of an experimenter; thus, they did not produce their own rhythmic movement. This suggests that the observed movement-auditory interactions are likely to originate in an aspect of the movement that does not involve motor planning. Further studies indicate that the vestibular system, which gives us our sense of balance and location in the gravitation field, is crucial to the influence of movement on hearing. Direct galvanic stimulation of the vestibular nerve on either every second or on every third beat of the ambiguous rhythm, such that people have the sensation that their head is moving from side to side in the absence of any actual movement, also influences whether people interpret the pattern as a march or as a waltz (Trainor et al. 2009). Interestingly, the vestibular system emerges very early in development (Romand 1992), and young infants love vestibular stimulation in the form of rocking, bouncing, and being moved energetically through the air. Therefore, vestibular input is prominent during the time when musical processing first emerges.

The ability to reproduce rhythmic patterns develops through the preschool years. However, the same organizational principles appear to apply to Western children and adults: duple meters are easier than triple, rhythms containing fewer different note durations are easier, and intensity accents delineating the metrical structure improve performance. Drake found that 7-year-olds were more accurate than 5-year-olds at reproducing short rhythms, but that musically untrained adults were no more accurate than 7-year-olds (Drake 1993). From the age of 4 years, children demonstrate the ability to extract metrical structure in that they can tap synchronously to a beat (Drake et al. 2000b). Drake also found that the ability to tap synchronously to rhythm patterns improves between 4 and 11 years of age (Drake et al. 2000b). With increasing age, children also improve in their ability to tap flexibly at faster and slower levels of the metrical hierarchy, suggesting an improvement in complex metrical processing. Interestingly, the preferred tapping tempo decreases with age and with musical training, suggesting that the ability to process longer time spans improves with age (Drake et al. 2000b; McAuley et al. 2006). Little is known about younger children as they do not readily do tapping tasks. In order to test younger children's ability to move synchronously to an auditory beat, Eerola et al. (2006) recorded and analyzed the movements made by 2- to 4-year-old children to music. Although many children hopped, circled, or swayed to the music, they did not show evidence of changing the tempo of their movements to match changes in the tempo of the music. The literature suggests, then, that young infants can perceive metrical structure in music and that they link it to, and are influenced by, proprioceptive cues to movement. However, producing movement that is coordinated with an external auditory rhythm is more difficult and takes years to master.

The fact that different musical systems use different scales and harmonic structures, and that enculturation to the pitch structure of one's culture takes place through simple exposure to the music of that culture was discussed in Sect. 4.2.1. Is the same true for metrical structure? The intimate connection between movement and music suggest that our sense of rhythm in music might originate in the regularities of movements such as heartbeats and locomotion. However, there is also a learned and culture-specific aspect of musical meter as well. Many movements, such as walking, involve simple 1:1 or 1:2 metrical structures, as in a march. In Western music, 1:1 and 1:2 ratios predominate (Fraisse 1982), and Western adults are better able to detect changes in rhythms with simple meters than rhythms with more complex meters (Hannon and Trehub 2005a; Repp et al. 2005; Snyder et al. 2006). Indeed, when confronted with complex meters such as groups of 5 or 11, Western listeners attempt to approximate them with simple meters (Hannon and Trehub 2005a).

Unlike Western music, many other musical systems use complex meters, even in their folk music. Adults who grew up listening to music with such complex rhythms have no trouble processing them, as evidenced in their ability to detect subtle timing changes, even in the absence of formal musical training. For example, Hannon and Trehub (2005a) showed that Bulgarian and Macedonian adults, whose folk music contains non-isochronous meters with ratios of successive beats being, for example, 2:3 (e.g., a group of two beats followed by a group of three beats), had no problem detecting duration changes in complex rhythms typical of their culture's music. On the other hand, North American listeners were unable to detect these changes, although they could readily do so with simple meters using 1:2 ratios.

Hannon and colleagues have shown that sensitivity to system-specific rhythmic structure develops during the second half of the first year after birth. Just as 6-month-olds do not yet process pitch structures according to the scale system of their culture, 6-month-olds do not yet process metrical structures according to the structures dominant in their culture. Hannon and Trehub (2005a) showed that 6-month-old Western infants could detect changes in rhythms with both simple meters typical of Western music and complex meters typical of Bulgarian and Macedonian folk music. However, by 12 months, Western infants had lost the ability to process rhythms with complex meters, indicating that they had become enculturated to the dominant metrical patterns of the music in their environment (Hannon and Trehub 2005b). Bergeson and Trehub (2006) further showed that by 9 months, infants more readily process duple (march-like) rhythms compared to triple (waltz-like) rhythms, again paralleling their experience, as duple meters predominate in Western musical structure.

In sum, for infants, as for adults, musical rhythm is intimately connected to rhythmic movements, and infants are precocious musical rhythm processors. The evidence suggests that infants' rhythm perception becomes specialized for the specific metrical structures of their culture by 1 year of age, although experience with a foreign meter at this age will reinstate their lost ability to process it (Hannon and Trehub 2005b). It is not clear how long this period of plasticity for different rhythm types persists. However, rhythm perception and production continues to be refined through childhood even in the absence of formal musical training (Drake et al. 2000b).

4.3.2 Differences Between Adult Musicians and Nonmusicians in Musical Rhythm Processing

Consistent with the idea discussed in the preceding text that rhythm is more fundamental to music than pitch, cases of impaired rhythm processing appear to be much more rare than cases of tone-deafness (Foxton et al. 2006). That said, there is probably a wide range of rhythmic abilities in the general population, as well as a wide range of abilities to dance. To some extent, these differences are likely the result of different amounts and types of musical training. At the present time, the extent to which formal musical instruction in childhood is necessary for the development of good rhythmic perception and production skills remains unknown. However, a few brain imaging studies show that adult musicians and nonmusicians differ in their responses to rhythm. It is not known, of course, whether these differences are the result of the training experience or the extent to which those individuals with innate predispositions for good rhythmic processing gravitate to music lessons. However, they are consistent with the notion that musical training affects brain development and subsequent competence for rhythmic processing. For example, Jongsma et al. (2004) measured ERPs while sequences of woodblock sounds that contained omissions of five beats in a row at unpredictable times were presented to musicians (drummers and bass guitarists) and nonmusicians. Subjects were to tap at the time of the fifth omitted sound. The musicians showed less variability in tapping on the omitted tone indicating a better ability to keep an accurate internal beat in the absence of sound. Consistent with this, in both groups, a positive slow wave response to omitted tones was seen that showed less latency jitter in the musician group.

There is evidence that, with musical training, rhythm processing shifts from a predominant focus in the right hemisphere to a predominant focus in the left hemisphere. fMRI studies of rhythm processing reveal greater left activation in musicians than in nonmusicians (Limb et al. 2006). Using MEG, Vuust et al. (2005) found that MMN responses to deviations in fairly complex rhythms were larger on the right in nonmusicians, but larger on the left in jazz musicians. They interpreted this finding as having to do with rhythm being a communication system (similar to language) used by jazz players who must synchronize while communicating musical ideas through their improvisations. However, it could also be that the left hemisphere is simply better at the precise interval timing that trained musicians develop (Zatorre 2001), whereas the right hemisphere is critical for ordinal timing related to sequencing. Rhythm processing activates a network of auditory and movement-related regions that are similar in musicians and nonmusicians (Limb et al. 2006). What appears to differ between groups is the degree of activation, suggesting that one of the effects of musical training is simply to recruit more neurons to the tasks of perceiving and producing musical rhythms. This is also consistent with behavioral data showing that musicians and nonmusicians approach rhythmic processing in qualitatively similar ways, but that musicians, in comparison to nonmusicians, simply perform better with familiar types of metrical structures that are compatible with their musical training (Jones et al. 1995; Jones and Yee 1997).

Recent models of rhythmic entrainment suggest that the brain acts like a bank of oscillators, each of which can be driven best with different frequencies of input (e.g., Large and Jones 1999). EEG and MEG data lend support to this idea in that brain responses to sound contain rhythmic oscillations that are affected by sound input. Sound events cause bursts of evoked gamma band activity (30–100 Hz) between about 50–100 ms after sound onset. Of more interest here is induced activity, which also occurs in response to a sound, but which is not precisely time-locked to the sound (e.g., Shahin et al. 2008). Induced gamma band is thought to represent recruitment of intrinsic rhythms that occur in the absence of sound to the processing of the sound. Snyder and Large (2005) showed that induced gamma band activity can occur even when there is no physical sound, if a rhythmic sequence of sounds sets up an expectation for a sound event at that time. Gamma band activity in response to isolated tones is larger in musicians than in nonmusicians for musical tones, and develops earlier in children taking music lessons than in those not training musically (Shahin et al. 2008). Bhattacharya et al. (2001) found that gamma band synchrony is greater in musicians than in nonmusicians when listening to music, but not when listening to text or in silence. This suggests that music may be more meaningful to musicians than to nonmusicians. Recent research suggests that there are interactions between different frequencies of oscillation in the brain. Fujioka et al. (2009) found that activity in the beta band (15–30 Hz), associated with the motor system, decreased after each tone in a rhythmic sequence and slowly increased thereafter. However, it did not decrease after a tone omission. In contrast, activity in the gamma band, associated with memory and attentional functions, increased after each tone, and increased after tone omissions as well. Little is known as of yet about how oscillatory activity in the brain develops and how it is affected by musical training at different ages. However, because oscillatory activity reflects the processing of sequences over time, it is an important area for future research.

In sum, there is evidence that although musicians and nonmusicians use similar networks in the brain to process rhythm, there are substantial differences in their brain responses to rhythms, with musicians' responses typically earlier and larger. There are also laterality effects, with musicians typically using the left hemisphere to a greater extent than nonmusicians. The processes involved in rhythm perception and how they differ between musicians and nonmusicians is only beginning to be understood. There is even less work on how these networks develop, whether there are sensitive periods for the development of rhythmic brain responses, and whether sophisticated musical rhythmic behavior can be taught in adulthood.

4.3.3 Effects of Formal Musical Training on Children's Processing of Rhythm

There is very little scientific literature on the effects of formal musical training on rhythm processing in children. Drake (1993) found that by 7 years of age, nonmusically trained children were similar to nonmusically trained adults in their

ability to reproduce heard binary and ternary rhythms. Adult musicians, on the other hand, were considerably better. In another study, Drake et al. (2000b) found that both child (6–10 years) and adult musicians were superior to age-matched nonmusicians at rhythmic reproduction, suggesting that musical training plays a large role in rhythmic competency. Tapping rate and flexibility to tap at different levels of the metrical hierarchy are also better in child musicians than nonmusicians. It remains largely unknown, however, the ages at which it is best to engage in musical training in order to achieve optimal rhythmic proficiency.

It is also possible that musical training might accelerate enculturation to rhythmic forms dominant in the musical system of exposure. Because rhythmic enculturation is seen in the second half of the year after birth (Hannon and Trehub 2005b; Hannon and Trainor 2007), it might be necessary to investigate effects of musical training by comparing the amount or type of musical experience infants receive. One study provides suggestive evidence that music classes can in fact accelerate acculturation to the dominant rhythms of a culture. Gerry et al. (2009) compared 7-month-old infants who were taking Kindermusik classes with infants who were not taking any infant music classes on the interaction between movement and interpretation of a metrically ambiguous rhythm. Specifically, they ran a group of Kindermusik infants through the same procedure as in Phillips-Silver and Trainor (2005), described in Sect. 4.3.1. Kindermusik classes involve parents walking, running, or otherwise moving their infants to the music on prescribed tapes. The tapes reflect Western rhythmic biases, with most of the songs being in duple meter and the rest in triple meter.

There were two interesting findings. First, on the preference test, infants in Kindermusik classes listened longer to the rhythm patterns on each trial than did infants not engaging in music classes, indicating a heightened interest in the rhythms. Second, the Kindermusik infants resembled nonmusically trained infants in that movement experience (duple, triple) to the metrically ambiguous rhythm resulted in a preference for an accented version of the rhythm that matched the movement experience (duple, triple). But only in the Kindermusik data was this effect stronger for duple bouncing than for triple bouncing. Thus, enriched experience with movement and music in duple time at 7 months of age appears to accelerate cultural biases for duple meter.

Of course, the end of the first year does not mark the end of the ability to learn complex meters without prior exposure. Using the same methods as in Hannon and Trehub (2005a), Hannon and Trehub (2005b) have shown that a little experience with complex rhythms at 12 months reinstates the ability to process complex meters. It currently remains unknown as to when training on particular rhythmic forms needs to occur in order to acquire native processing of these forms. However, by adulthood it appears that complex rhythms are difficult to learn without prior exposure.

4.3.4 Summary of the Acquisition of Musical Rhythm

Young infants are precocious at processing musical rhythm in that they can discriminate rhythmic patterns and connect rhythmic movement with their perception of auditory rhythms. At the same time, their processing of musical rhythm becomes

specialized for the rhythms in their environment between 6 and 12 months of age. Although there is little research on children, it appears that rhythmic processing for culturally relevant rhythmic structures improves during childhood. Children acquire musical rhythmic knowledge without formal training; still, compared to nonmusicians, adult musicians are better able to perceive rhythmic hierarchies and show enhanced brain responses and greater left laterality when processing musical rhythms. It remains unknown as to whether there is a sensitive period for musical rhythm acquisition, but it appears difficult for adults to perceive and produce complex rhythmic patterns if they have not been exposed to them during childhood.

4.4 Musical Emotion

For most of the population, music can elicit powerful emotions. Music appears to have the ability to lift the spirits, calm the nerves and express deep sorrow. Meyer (1956) proposed that although music does not generally refer to anything outside of the music itself, violation of musical expectations can elicit powerful emotions. Huron (2006) has expanded this idea in a theory of musical expression. Physiological responses to musical emotion confirm that music does activate limbic and cortical brain structures associated with emotion (Blood and Zatorre 2001; Peretz 2001; Schmidt and Trainor 2001). Further, people report physiological responses to music such as shivers down the spine, lump in the throat, laughter, and tears (Sloboda 1991). Direct measures of autonomic responses reveal that music affects breathing rate, heart rate, heart rate variability, and finger temperature (Nyklíček et al. 1997; Krumhansl 1997), although the exact relation between these responses and the specific emotion experienced remains unclear. Still, different ideas exist as to the nature of musical emotion. These ideas are explored in detail in this volume (see Hunter and Schellenberg, Chap. 5). Here we focus on questions related to the development of musical emotions; however, theories about musical development can also inform the debate as to the nature of emotional expression in music and why it evolved. For example, one major question concerns the evolution of emotional expression through music. Trainor and Schmidt (2003) suggested that emotional expression in music may have originated in infant-directed singing, serving the purpose of strengthening the emotional bond between preverbal infants and their caretakers. Another major question for research on musical emotion concerns the extent to which emotional responses to musical features are innate and universal or learned through exposure to a particular musical system (Kivy 1980). Some musical features appear to have similar emotional effects across cultures, such as a slow tempo for sad music and a fast tempo for happy music (see Hunter and Schellenberg, Chap. 5). However, other features, such as the use of the major mode in Western music to convey happiness, appear to require familiarity with particular musical conventions. In the following sections, we explore the development of emotional responses to music in the context of cross-cultural evidence.

4.4.1 *The Perception of Emotion in Music*

Adults report that emotional regulation, emotional induction, and enjoyment are the main motivational factors behind music listening (Juslin and Laukka 2004). For prelinguistic infants, emotional responses to music appear to be central as well. Many studies suggest that infant-directed speech and singing communicate emotion, regulate mood, and promote child–parent bonding (Trainor 1996; Trehub and Trainor 1998; Dissanayake 2000). Infant-directed speech is often referred to as “musical” or “emotional” speech, and like infant-directed singing, it differs from adult-directed communication in a number of characteristics including higher overall pitch, slower rate and exaggerated pitch contours (Fernald 1991; Papoušek 1992; Trainor et al. 1997). Infants prefer listening to infant-directed over adult-directed speech (Werker and McLeod 1989; Cooper and Aslin 1990) and singing (Trainor 1996). Further, infants appear to be more engaged by maternal singing than maternal speech (Nakata and Trehub 2004). Indeed, the purpose of these infant-directed modifications seems to be the expression of emotion and elicitation of attention (Fernald 1993; Trainor et al. 1997; Rock et al. 1999; Trainor et al. 2000). Using music for emotional regulation appears to be a human universal, as mothers across cultures sing to their infants to calm them or lull them to sleep (Trehub and Trainor 1993; Trehub et al. 1993a,b). Thus, early in development, caregivers use music in its simplest form as a means to communicate and bond emotionally with a child who has not yet developed language.

Although infants are sensitive to the emotional aspects of infant-directed speech and singing, they are not as sophisticated as adults at perceiving emotion in music, especially in instrumental music. With regards to musical structure, there appear to be two general types of cues to musical emotion: (1) basic acoustic cues that likely have their roots in the nonverbal vocal expression of emotion and (2) learned, musical-system-specific cues that have no obvious counterpart in vocal expression (Balkwill and Thompson 1999). The basic acoustic cues are understood early in development and simply become refined with experience, and appear to be universal across musical systems, while the musical-system-specific cues develop through exposure to the music of one’s culture. Each of these is discussed in the subsequent sections.

4.4.1.1 **Basic Acoustic Cues**

Some cues to emotional expression in music are shared with cues to vocal expression of emotion (i.e., prosody). Scherer’s (1985, 1986) component process model of affective states proposed that different emotions have particular effects on the somatic nervous system, which ultimately affects the voice. Based on this model, Scherer made predictions about the acoustic cues that accompany five basic emotions (anger, disgust, fear, happiness, and sadness). These cues primarily consist of features common to both music and language such as tempo or speech rate, loudness, and timbre.

Juslin and Laukka (2003) tested these predictions by conducting a meta-analysis of 104 studies of vocal expression and 41 studies of music performance. Their primary goals were to examine whether certain basic emotions (i.e., anger, fear, happiness, sadness, and love-tenderness) could be communicated to adult listeners and to profile the patterns of acoustic cues that accompany each of these emotions. The results suggested that adults were highly accurate at identifying basic emotions expressed by the prosody of the voice both within- and cross-culturally, as well as by vocal and instrumental musical performance. In line with these findings, Juslin and Laukka found a significant degree of overlap between the kinds of cues used to express emotion vocally and musically. However, they stress that individual acoustic cues to emotion are neither necessary nor sufficient to identify specific emotions, but rather, are combined in an additive fashion to form patterns that probabilistically indicate certain emotions. The authors posit that expression in music essentially mimics vocal expression of emotion, and that these shared, domain-general acoustic cues are responsible for the bulk of the emotional message conveyed in music.

Support for the use of basic acoustic cues to emotion in music comes from cross-cultural studies of emotion perception. Balkwill and Thompson (1999), for example, found that Western listeners could identify the intended emotions of joy, sadness, and anger in Hindustani ragas of classical Indian music, and that they based their judgments on acoustic cues such as tempo, melodic and rhythmic complexity, and timbre. Balkwill et al. (2004) later extended this work to show that Japanese listeners could accurately identify the same emotions from Japanese, Western, and Hindustani musical excerpts. Again, listeners based their judgments on acoustic cues such as tempo, melodic and rhythmic complexity, and timbre, as well as loudness; however, in support of Juslin and Laukka (2003), Balkwill et al. (2004) found that these adult participants used combinations of cues to identify emotions rather than relying on individual components. Furthermore, Adachi et al. (2004) found that Japanese 8- to 10-year-olds and adults successfully identified whether Canadian 8- to 10-year-olds were intending to convey happiness or sadness through singing. Thus, it appears that certain combinations of acoustic cues are reliably used across musical systems to convey particular emotions, although future research should examine the use of these cues in a wider variety of musical systems.

The developmental question of interest is whether sensitivity to these universal acoustic cues to emotional expression in music is present early in development and whether this sensitivity is affected by maturation and experience. Unfortunately, only a few studies have examined emotional reactions to music in infancy. Rock et al. (1999) found that adults could identify whether 6- to 7-month-olds were listening to a lullaby or a play song based on the infants' behavior during each kind of song. Infants tended to focus inward, looking down, while listening to lullabies and on their caregivers when listening to play songs, suggesting that they associated lullabies with calming and sleep, and play songs with arousal and activity. In an attempt to show that infants can link emotions conveyed through music to those conveyed through visual means, Nawrot (2003) found that 5- to 9-month-olds looked longer at an

emotionally concordant dynamic visual display (i.e., happy visual images) during happy music but not at sad visual images during sad music. The lack of preference for the concordant display during sad music may have reflected the infants' inability to recognize emotions conveyed by music. In particular, classical pieces composed by Mozart and Beethoven were used as stimuli, and these, in combination with the dynamic nature of the visual displays, may have been too complicated for especially the younger infants to demonstrate their knowledge (see also Schmidt et al. 2003). However, another possibility is that infants' lack of preference for the concordant display during sad music reflected infants' avoidance of engaging in sad displays of emotion in general. Schmidt et al. (2003) examined EEG responses in infants at 3, 6, 9 and 12 months of age to classical musical excerpts expressing joy, sadness and fear. Unlike with adults, who show more left activation for positive emotions and more right activation for negative emotions (Schmidt and Trainor 2001), infants' EEG responses did not differentiate the excerpts by emotional valence. It is possible that infants do not show different emotional responses to music, but it is also possible the musical excerpts were simply too complex for infants to extract emotional meaning. Of most interest, across all excerpts, music generally increased brain activity at 3 months, had little effect at 6 and 9 months, but decreased brain activity at 12 months. This pattern suggests that music has an arousing effect in early infancy, but a calming effect in later infancy.

Although we know little about emotional responses to music in infancy, there is somewhat more research with preschool-aged and school-aged children. Trainor and Trehub (1992b) showed that children as young as 4 years of age can associate music with non-musical referents in that they can choose the correct animal picture to go with an excerpt from Prokofiev's *Peter and the Wolf* and Saint Saen's *Carnival of the Animals*. Further, when asked why they made their choices, they sometimes cited the emotion expressed in the music (e.g., the wolf music was scary). Children as old as five appear to rely heavily on tempo (fast = happy, slow = sad) in their emotional judgments of music (Dalla Bella et al. 2001). Similarly, 6- to 12-year-olds largely base their judgments of happiness and sadness on rhythmic activity and articulation, and their judgments of excitement and calm on rhythmic activity and meter (Kratus 1993). Further, the development of happiness, sadness, and anger recognition in musical pieces appears to parallel the recognition of those same emotions in vocal intonation but not facial stimuli in 3- to 12-year-olds (Brosgole and Weisman 1995). A final point of interest is that when asked to sing a familiar song to convey happiness or sadness, 4- to 12-year-olds make use of domain-general cues such as tempo, loudness, and voice quality but tend to ignore music-specific cues such as the use of *legato* (singing smoothly as opposed to choppy), especially if they are younger and less experienced with music.

Taken together, the evidence suggests that vocal cues to emotional expression are heavily exploited even in instrumental music, and that adults and children can use these cues to infer the intended emotion being expressed. This may explain why adults and even children can identify musical emotions cross-culturally, despite being unfamiliar with the musical structure of that culture. While more research is needed to reveal whether infants can make use of some of these cues and how this

ability develops, the available evidence suggests that children are able to make use of some basic acoustic cues to emotional expression such as tempo, rhythmic activity, and loudness.

4.4.1.2 Musical-System-Specific Cues

Despite the finding that basic emotions can be communicated quite readily through cross-culturally universal acoustic cues, culture-specific cues can also convey various emotions. For example, Sloboda (1991) outlined the emotional-physiological reactions that tend to occur in response to a number of different structural features in music through a survey of experienced listeners and found that tears were largely evoked by melodic appoggiaturas (where an expected stable tone on a strong beat is delayed by a nonstable tone in a melodic line), whereas shivers down the spine tended to be evoked to unexpected harmonies in Western music. Another primary example in Western tonal music is the use of the major mode to convey happiness and the minor mode to convey sadness (Gabrielsson and Juslin 2003). Indeed, adult listeners readily make these mode-emotion associations (Hevner 1936; Husain et al. 2002; Gagnon and Peretz 2003).

Developmental research has confirmed that the major-happy minor-sad distinction is acquired during childhood even in the absence of formal musical training. Despite showing a preference for highly consonant over highly dissonant chords, 6-month-olds did not prefer major over minor chords, even though major chords are considered to be more consonant than minor chords (Crowder et al. 1991). Thus, infants are not yet sensitive to the emotional connotations of major and minor chords, unlike both musically trained and untrained adults. Interestingly, Kastner and Crowder (1990) tested 3- to 12-year-olds and found that even the 3-year-olds could associate major melodies with positive emotions and minor melodies with sad emotions, although this ability improved somewhat with age. However, most studies have failed to find any evidence for this association until approximately 6–7 years. For example, Gregory et al. (1996) replicated Kastner and Crowder's study with 3- and 4-year-olds, 7- and 8-year-olds and adults with one modification: instead of providing listeners with four response options as in the original study (schematic faces depicting happy, contented, sad, and angry), they simply provided two (schematic faces depicting happy and sad). The researchers found that 7- and 8-year-olds as well as adults paired happy faces with excerpts in the major mode and sad faces with excerpts in the minor mode; however, 3- to 4-year-olds failed to use mode to classify musical excerpts as happy or sad. Similarly, Gerardi and Gerken (1995) found that 8-year-olds and adults were sensitive to mode in their judgments of happy and sad melodies whereas 5-year-olds performed at chance levels. Another interesting finding from this study was that only adults associated an ascending melodic contour with happiness and a descending melodic contour with sadness, suggesting that this kind of association also develops during childhood. Finally, Dalla Bella et al. (2001) conducted a study that

allowed them to examine the relative effects of tempo and mode manipulations on happy-sad affective judgments in Western classical orchestral music. Adults' and 6- to 8-year-olds' judgments were sensitive to both mode and tempo manipulations, whereas 5-year-olds' judgments were only affected by tempo changes; 3- and 4-year-olds, in contrast, could not identify whether excerpts were happy or sad. These results suggest that by 5 years, tempo figures prominently in children's judgment of emotion in music, and that soon after mode also becomes an important factor.

In sum, although basic acoustic cues may predominate as the most salient cues to emotion in music, culture-specific cues can also be used in judgments of emotional expression. In Western music, the most important and widely researched of these cues is that of mode. Research has revealed that the major-happy and minor-sad associations are not yet present in infancy, and that they develop by the age of 6 or 7 years. Younger children tend to rely exclusively on basic acoustic cues such as tempo. Thus, as children acquire sensitivity to the structure of the music of their culture, so too do they acquire sensitivity to the cues to emotional expression that are based on this musical structure.

4.4.2 The Effects of Experience on the Perception of Emotion in Music

In the context of lyricized songs, evidence suggests that children as old as 10 years rely more heavily on the emotional connotations of lyrics than on particular musical features (Morton and Trehub 2007). However, other studies have examined children's perception of emotion in instrumental music on the assumption that children can base their emotional judgments on multiple musical features in the absence of accompanying words. These studies have generally found small age-related differences between 4 and 12 years of age in identification of happy and sad emotions in music. There is some evidence of a difficulty identifying angry and fearful emotions especially in younger children, a difficulty that is also reflected in adults' lower agreement on their judgments of these two emotions in music (e.g., Cunningham and Sterling 1988; Dolgin and Adelson 1990; Terwogt and van Grinsven 1991; Kratus 1993; Robazza et al. 1994). In addition, valence (positive or negative) appears to be more salient than intensity (high or low) when it comes to children's musical emotional perception (Kratus 1993) (see Hunter and Schellenberg, Chap. 5, for a discussion of models of emotional response to music). It therefore appears that children as young as four can identify basic emotions expressed by music when multiple cues to emotion are involved. It remains unknown whether immaturities in recognizing emotions in music are a function of general immaturities in emotion recognition or specific to musical materials.

Generally, musical training appears to have little effect on emotion perception in music. This is evident, at least, with regard to recognition of basic emotions in

adults (Hevner 1935; Edmonston 1966). Similarly, neither has musical training been found to affect emotion recognition in children (Terwogt and van Grinsven 1991; Robazza et al. 1994). Sloboda (1985) has suggested that effects of musical training play a role only in the identification of more subtle or more secondary emotions conveyed in music, which may explain why most studies that examine basic emotion recognition fail to find any effect of musical experience. This argument may extend to developmental studies that also tend to focus on basic emotions – it seems likely that more age-related differences would emerge in the identification of subtle or nonbasic emotions in music. At least one study has found a musical expertise advantage in a related domain, however. Thompson et al. (2004) conducted a series of experiments examining the identification of happiness, sadness, fear, and anger from speech prosody and from tone sequences that mimicked speech prosody. Adult musicians outperformed nonmusicians on emotion identification, especially those of sadness and fear. Seven-year-olds who had been randomly assigned initially to keyboard lessons and had recently completed 1 year of musical training outperformed nonmusicians of the same age on identifications of anger and fear. As mentioned above, these also tend to be the emotions that are easily confused in music.

Taken together, it appears that musical training may influence the subtleties of emotion perception in speech and music, but that even untrained listeners are highly sensitive to the primary message being conveyed.

4.4.3 Summary of Development of Musical Emotion

The study of the development of musical emotion is in its infancy and, as such, many questions remain for future research to address. Research on children's perception of emotion in music has revealed that even children as young as 4 years can recognize emotions conveyed by music, and that they largely use acoustic cues that pertain to emotional expression in both speech and music. However, little is known about musical emotion perception in infants and toddlers, an area that deserves empirical attention. In addition, children also learn to use culture-specific cues such as mode to identify emotions conveyed by music. Future research should examine the developmental trajectories for other musical-system-specific cues.

Most studies have failed to find a significant effect of musical training on the perception of musical emotion, even in children. However, these studies have tended to examine the perception of basic and highly agreed-upon emotions in music. Research should therefore examine whether music lessons in childhood lead to a greater sensitivity to more subtly expressed emotions in music. Another question that remains open is whether early musical training might accelerate the acquisition of learned cues to emotional expression. By examining how the perception of musical emotion is shaped through developmental as well as musical experience, research can shed light on the universal appeal of music.

4.5 The Effects of Musical Training on Other Cognitive Skills

Over the past few decades, research on the possible cognitive, social, and emotional benefits of musical training has received a great deal of attention from educators, clinicians, and the popular media. Many claims have been made about music's ability to promote overall well-being, to improve performance in mathematics, spatial-temporal abilities, language, and memory, and of course, to generally "make you smarter." Are these claims justified? Unfortunately, much of the research investigating the benefits of music lessons has methodological difficulties that preclude the inference of causation. For example, most research has implemented correlational or quasi-experimental designs (i.e., without random assignment to experimental conditions), leaving open the possibility that any differences found between musicians and nonmusicians could be accounted for by preexisting differences between those who chose to take music lessons and those who did not. In addition, many studies have failed to include an adequate control group, simply comparing those who received musical training with those who received no extracurricular stimulation. Differences between these groups could be the result of the additional individual attention received by the music group, or the school-like nature of musical training, and differences might not be caused by the *musical* training in particular (see Schellenberg 2001 for a review of the methodological difficulties of musical training studies). Despite these problems, mounting evidence suggests that musical training does lead to small but reliable gains in mathematics, spatial-temporal abilities, language, and general intelligence. Each of these will be reviewed in subsequent sections, followed by a discussion of whether memory and attention might mediate these effects.

4.5.1 *Mathematical Abilities*

Despite anecdotal reports of the link between music and mathematics, in fact, little research has been conducted on the topic. That music and arithmetic may involve similar processes is not surprising given the abundance of mathematical relations in the harmonic, rhythmic and metrical aspects of music. A few correlational studies without random assignment have examined whether musicians tend to score higher on mathematical achievement tests than individuals who do not participate in any musical training. For example, Gouzouasis et al. (2007) found that high school students who participated in music classes tended to score higher on standardized tests of academic achievement, especially mathematics, and Cheek and Smith (1999) observed this effect in middle-school students who had received at least 2 years of private music lessons. These results support the link between music and math, but do not shed light on the direction of causation. Gardiner et al. (1996) attempted to address this question by testing 5- to 7-year-olds involved in a visual arts and music curriculum compared to students not in

special classes. The researchers found that after 1 and 2 years of participation in the program, children in the experimental arts program outperformed controls on tests of mathematical skills, despite the finding that these children initially scored lower on mathematics tests than control children. However, since children in the experimental group also received visual arts training, musical training cannot be isolated as the driving factor behind the improvement.

Finally, Vaughn (2000) conducted two meta-analyses on 20 correlational and five experimental studies on the relationship between musical and mathematical abilities. The results of the analysis on correlational studies suggest that adult musicians tend to score higher on mathematics achievement tests than nonmusicians. Similarly, the meta-analysis of children's performance suggests that involvement in musical training leads to improvements in mathematical performance. However, as noted by the author, very few of the studies included in the meta-analyses came from peer-reviewed journals, suggesting a general lack of well-conducted studies examining the relationship between music and mathematics. In line with this observation, several peer-reviewed studies have failed to find support for the link between musical training and mathematical performance (e.g., Costa-Giomi 1999, 2004; Forgeard et al. 2008). Taken together, it appears that support is mixed for this relationship, and more research on the topic is warranted. It is possible that the inconsistencies arise because musical training affects some aspects of mathematical performance and not others. Thus, it would be useful for future studies to make a distinction between the different subskills (e.g., arithmetic, trigonometry, algebra, calculus) that are involved in mathematical ability as opposed to simply measuring mathematical achievement through grades achieved in school.

4.5.2 Spatial–Temporal Abilities

Somewhat related to mathematical skill is the ability to visualize and mentally manipulate images in space and time. The relation between this ability and music has received a good deal of attention, stemming primarily from the widely publicized “Mozart Effect” in which music listening is proposed to lead to a short-term improvement in spatial-temporal performance (Rauscher et al. 1993). Although this effect has been shown to be short-lived, somewhat unreliable, and primarily accounted for by arousal (see Schellenberg 2001, 2005 for reviews), stronger evidence exists for an improvement in spatial–temporal ability following musical training. Hetland (2000) conducted a meta-analysis of 15 studies examining the effects of musical training on spatial–temporal abilities in 3- to 12-year-olds, and another meta-analysis of eight studies on the effects of music instruction on a wider range of spatial abilities. The results of both analyses suggest that musical training does lead to significant and reliable gains in spatial reasoning. Further analyses suggest that the largest gains are found for younger children, individual as opposed to group lessons, and programs teaching standard musical notation; however, the author stressed that significant improvement was seen for all ages and all instruction types.

Hetland did not find support for several other variables that were predicted to mediate effect sizes, including socioeconomic status, length of musical training, parental involvement in music lessons, keyboard instruction, implementation of expressive movement in lessons, and inclusion of composition and/or improvisation in lessons. Further, the results could not be accounted for by experimenter bias, nonspecific effects of extracurricular musical training, preexisting differences between musically trained and untrained children, or study quality. Although the results are strong, more research is needed to verify each of the variables suspected of mediating the effect. In a similar argument to that made above in Sect. 4.5.1, Hetland called on future research to pinpoint whether certain kinds of spatial abilities are improved after musical training, or if the effect generalizes to all spatial tasks. Further, some evidence suggests that musical training may initially accelerate the development of spatial abilities, but that the advantage may later disappear as musically untrained children catch up in ability (Costa-Giomi 1999). In any case, most research seems to point to a reliable effect of musical training on spatial-temporal skills in children.

4.5.3 *Language Skills*

Despite correlations between musical ability and early reading skill (Anvari et al. 2002), and evidence that musical and linguistic processing share neural resources (see Patel 2008), few studies have specifically examined the effect of musical training on language skills. However, some evidence suggests that participation in musical activities leads to improvement in several linguistic abilities including linguistic pitch processing, phonological awareness, and early reading ability. For example, research suggests that musical training leads to superior linguistic pitch processing in both adults (Schön et al. 2004; Marques et al. 2007) and children (Magne et al. 2006; Moreno and Besson 2006; Moreno et al. 2008). These results may be linked to the finding that musically trained children and adults are better at processing emotional meaning from speech prosody than are untrained adults and children (Thompson et al. 2004).

Phonological awareness includes the discrimination, detection, and manipulation of linguistic sounds and has been found to be highly predictive of early reading ability (see Bus and van IJzendoorn 1999). Since this skill relies heavily on basic auditory processing, it may be unsurprising that it appears to benefit from musical training. Gromko (2005) found that after 4 months of music classes during regular kindergarten classes, children performed better than controls on a phonological awareness task. Similarly, Overy (2003) found that musical training improved phonological awareness in children with dyslexia. Further research should replicate these claims, but it is reasonable to conclude that phonological skills can benefit from musical training.

The relationship between musical training and reading is less clear. Butzlaff (2000) conducted a meta-analysis on six experimental or quasi-experimental studies

to investigate whether musical education affected reading, and found a small and unreliable mean effect size. However, it would be premature to conclude that musical training has no effect on reading skills because very few studies were included in this analysis, and because each study used measures that assess different aspects of reading, such as word decoding (identifying words based on their graphemic representation) and reading comprehension. The only published study in this analysis examining word decoding ability as the outcome of interest found a strong effect of musical training in a group of poor readers (Douglas and Willatts 1994). By contrast, other studies have failed to find an effect of music education on word decoding ability (Overly 2003; Gromko 2005); however, the length of training in these studies was shorter than the 6 months of training reported by Douglas and Willatts. Thus, the apparent discrepancy in the results of studies examining reading benefits following formal music lessons probably stems from several reasons: (1) musical training is likely to be more closely tied to word decoding ability than to reading comprehension since both of the former involve mapping abstract visual symbols to sounds; (2) if this first point is true, there may be an optimal time in development (e.g., when children are first learning to read) where musical training exerts maximal benefits on word decoding, before which or beyond which the effects are likely to be reduced; (3) since reading is a complex and multi-faceted skill, longer training may be necessary to produce effects on reading in contrast to the case for phonological awareness, a precursor to reading; and (4) musical training may affect the reading abilities of normal and poor or dyslexic readers differently. Clearly, more research is needed to elucidate the relationship between musical training and reading skills.

4.5.4 *General Intelligence*

As has been reviewed in the preceding text, there appears to be evidence for both non-verbal and verbal benefits following musical training. This leaves open the possibility that many of these relationships may be explained by gains in general intelligence rather than specific transfer effects. Indeed, several studies have found a link between musical training and general intelligence (e.g., Costa-Giomi 1999; Bilhartz et al. 2000; Forgeard et al. 2008). In a strong test of this hypothesis, Schellenberg (2004) conducted a seminal study examining the effect of musical training on intelligence in 6-year-olds. Children were randomly assigned to 1 year of keyboard lessons, singing lessons, drama lessons or no lessons at no cost to their families (the last group received lessons the following year). Children were administered a standardized IQ test (the Wechsler Intelligence Scale for Children-III, or WISC-III) before beginning lessons and again after a year of lessons. Schellenberg found that while all children showed increases in full-scale IQ, children in the music groups (keyboard and singing lessons) improved more than children in the control groups (drama lessons and no lessons). Further, the greater gains in IQ experienced by the music groups were not limited to particular areas assessed by the WISC-III

such as verbal comprehension, but rather, these gains were present across the different subtests and intellectual areas of the WISC-III. Schellenberg therefore concluded that musical training led to small but reliable gains in full-scale intelligence. As these children participated in group lessons, it is possible that individual lessons would lead to even larger effects on intelligence, since this type of relationship has been found for a subcomponent of intelligence, namely, spatial-temporal reasoning (Hetland 2000). In addition, Schellenberg (2006) later showed that the duration of music lessons in childhood was correlated with general intelligence in 6- to 11-year-old children as well as in undergraduate students, and Forgeard et al. (2008) later replicated these findings with 8- to 11-year-olds. It should be noted, however, that the results of these studies stand in contrast to those of Costa-Giomi (1999) who reported that the cognitive benefits following musical training might be short lived. Future research needs to clarify the long-term effects of musical training and to include more studies with random assignment, but the evidence thus far supports the link between musical training and general intelligence.

4.5.5 Memory and Attention

If musical training has a general effect on cognitive processing, then the mechanisms of transfer likely involve improvements in general processing mechanisms. Two prime candidates are memory and attention. Active participation in musical activities involves many different learned skills, many of which can be linked to memory and attention. For example, much of music training involves memorization of auditory patterns, visual patterns (musical notation), auditory-motor sequences such as melodies and fingering patterns, and associated verbal labels. Learning an instrument requires attending to and assessing the sounds one is producing and modifying them to match an internal model. Similarly, group performance requires musicians to simultaneously attend to musical notation as well as their own and others' sounds, matching tempo, dynamics and style while ignoring other irrelevant sources of information. Further, music-making often requires sustained attention to one activity for long periods of time. Thus, it is reasonable to conclude that music lessons are likely to actively train memory and attentional systems.

Evidence for verbal memory improvement following musical training is strong. In one retrospective study, Chan et al. (1998) compared adults who had received at least 6 years of musical training in childhood and adults who had not received any musical training. The groups were matched on age, grade point average, and years of education. The researchers found evidence of enhanced verbal memory in the musically trained group, but no differences between groups on visual memory. Similarly, Ho et al. (2003) found enhancement of verbal but not visual memory in children who had participated in music lessons compared to controls. By contrast, Jakobson et al. (2008) found that highly trained musicians exhibited better immediate and delayed recall for both verbal and visual material. Another study found superior

verbal and non-verbal working memory in musically trained children compared to controls, although musically trained adults in this study only exhibited superior verbal short-term memory (Lee et al. 2007). Several mechanisms have been suggested to account for musicians' enhanced memory abilities. Jakobson et al. (2003) found a correlation between years of musical training and performance on verbal recall, an effect that they proposed was mediated by musicians' superior auditory temporal-order processing abilities. Similarly, Franklin et al. (2008) found that musicians outperformed nonmusicians on tests of long-term verbal memory, and suggested that musicians' advantage stemmed from superior verbal rehearsal mechanisms. Although more research is needed to understand exactly how memory is affected, memory is a strong candidate for the mechanism mediating general intelligence improvements following musical training.

A smaller body of research suggests that attention is also a candidate mechanism behind musical-training induced cognitive benefits. Scott (1992) found that preschoolers enrolled in music lessons performed significantly better on a vigilance-like attention task than control children who were either: (1) not involved in any organized classes or preschool, (2) enrolled in preschool, or (3) enrolled in both preschool and creative movement classes. Further, musically trained children in this study persevered longer on a task requiring them to replicate a complex block model, and they made fewer design errors than control children. Of course, causation is not clear in these studies because random assignment was not used. Fujioka et al. (2006) provided stronger evidence for a causal link in their finding that an MEG component (N2m) associated with memory and attentional processing changed more over the course of a year in children participating in music lessons than in children not participating. Similarly, Bialystok and DePape (2009) found that musically trained adults exhibited superior executive functions than untrained adults. In sum, musical training may well enhance attention, which in turn benefits cognitive development in general. Future research should examine the different aspects of attention that may be enhanced through participation in musical activities.

4.5.6 Conclusions About Benefits of Musical Training in Other Cognitive Domains

Although practical considerations often preclude research from investigating the link between musical training and other cognitive skills with experimental designs (e.g., with true random assignment), a large body of research suggests that the cognitive benefits of musical training are widespread. This is not to say that preexisting differences do not exist between children who go on to take music lessons and those who do not – in fact, it is highly likely that there are many of these such as socioeconomic status, parental education, and family involvement in musical activities. However, it appears that the benefits of musical training can in some cases go above and beyond these initial differences. As music is a multifaceted and

complex set of skills involving perception, cognition, and emotional expression as well as individual practice and group performance, it is perhaps not surprising that extensive musical training can benefit other domains.

4.6 General Conclusions

Musical acquisition has a very long developmental trajectory. Sensitivity to some aspects of music can be seen in early infancy, such as sensitivity to consonance, relative pitch, and discrimination of simple rhythm patterns. Other aspects, such as sensitivity to harmony, do not reach adult levels until well into the teenage years. The developmental trajectory and the level of musical competence achieved depend heavily on musical experiences. Everyday exposure to a particular musical system cultivates specialized processing of the musical structure of that system. Formal musical training appears to accelerate musical development, cultivate specialization, and affect the development of networks in the brain for processing music, with the end result in adulthood of greater perceptual skills and performance ability. Formal musical training also affects executive functions such as memory, attention, and inhibition, and leads to benefits across a wide range of cognitive processes.

Acknowledgments The writing of this chapter was supported by grants from the Natural Sciences and Engineering Research Council of Canada, the Canadian Institutes of Health Research, and the Grammy Foundation. We thank Terri Lewis and Andrea Unrau for insightful comments on an earlier draft.

References

- Adachi M, Trehub SE, Abe J (2004) Perceiving emotion in children's songs across age and culture. *Jpn Psychol Res* 46:322–336.
- Anvari SH, Trainor LJ, Woodside J, Levy BA (2002) Relations among musical skills, phonological processing and early reading ability in preschool children. *J Exp Child Psychol* 83:111–130.
- Balkwill L, Thompson WF (1999) A cross-cultural investigation of the perception of emotion in music: psychophysical and cultural cues. *Music Percept* 17:43–64.
- Balkwill L, Thompson WF, Matsunaga R (2004) Recognition of emotion in Japanese, Western, and Hindustani music by Japanese listeners. *Jpn Psychol Res* 46:337–349.
- Bangert M, Schlaug G (2006) Specialization of the specialized in features of external human brain morphology. *Eur J Neurosci* 24:1832–1834.
- Baruch C, Drake C (1997) Tempo discrimination in infants. *Infant Behav Dev* 20:573–577.
- Bergeson TR, Trehub SE (2006) Infants' perception of rhythmic patterns. *Music Percept* 23:345–360.
- Bermudez P, Lerch, JP, Evans AC, Zatorre RJ (2008) Neuroanatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. *Cereb Cortex* 2008; doi: 10.1093/cercor/bhn196.
- Bharucha JJ, Stoeckig K (1986) Reaction time and musical expectancy: priming of chords. *J Exp Psychol Hum Percept Perform* 12:403–410.

- Bharucha JJ, Stoeckig K (1987) Priming of chords: spreading activation or overlapping frequency spectra? *Percept Psychophys* 41:519–524.
- Bhattacharya J, Petsche H, Pereda E (2001) Long-range synchrony in the γ band: role in music perception. *J Neurosci* 21:6329–6337.
- Bialystok E, DePape A (2009) Musical expertise, bilingualism, and executive functioning. *J Exp Psychol Hum Percept Perform* 35:565–574.
- Bigand E, Pineau M (1997) Global context effects on musical expectancy. *Percept Psychophys* 59:1098–1107.
- Bigand E, Poulin-Charronnat B (2006) Are we “experienced listeners”? A review of the musical capacities that do not depend on formal musical training. *Cognition* 100:100–130.
- 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.
- Bilhartz TD, Bruhn RA, Olson JE (2000) The effect of early music training on child cognitive development. *J Appl Dev Psychol* 20:615–636.
- Blood AJ, Zatorre RJ (2001) Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proc Natl Acad Sci USA* 98:11818–11823.
- Bosnyak DJ, Eaton RA, Roberts LE (2004) Distributed auditory cortical representations are modified when nonmusicians are trained at pitch discrimination with 40 Hz amplitude modulated tones. *Cereb Cortex* 14:1088–1099.
- Brosigle L, Weisman J (1995) Mood recognition across the ages. *Int J Neurosci* 82:169–189.
- Bus AG, van IJzendoorn MH (1999) Phonological awareness and early reading: a meta-analysis of experimental training studies. *J Educ Psychol* 91:403–414.
- Butzlaff R (2000) Can music be used to teach reading? *J Aesthetic Educ* 34:167–178.
- Čeponienė R, Kushnerenko E, Fellman V, Renlund M, Suominen K, Näätänen R (2002) Event-related potential features indexing central auditory discrimination by newborns. *Cogn Brain Res* 13:101–113.
- Chan AS, Ho Y, Cheung M (1998) Music training improves verbal memory. *Nature* 396:128.
- Chang H, Trehub SE (1977) Infants’ perception of temporal grouping in auditory patterns. *Child Dev* 48:1666–1670.
- Cheek JM, Smith LR (1999) Music training and mathematics achievement. *Adolescence* 34:759–761.
- Clarkson MG, Clifton RK (1995) Infants’ pitch perception: inharmonic tonal complexes. *J Acoust Soc Am* 98:1372–1379.
- Cooper RP, Aslin RN (1990) Preference for infant-directed speech in the first month after birth. *Child Dev* 61:1584–1595.
- Corrigan KA, Trainor LJ (2009) Effects of musical training on key and harmony perception. *Ann N Y Acad Sci* 1169:164–168.
- Costa-Giomi E (1999) The effects of three years of piano instruction on children’s cognitive development. *J Res Music Educ* 47:198–212.
- Costa-Giomi E (2003) Young children’s harmonic perception. *Ann N Y Acad Sci* 999:477–484.
- Costa-Giomi E (2004) Effects of three years of piano instruction on children’s academic achievement, school performance and self-esteem. *Psychol Music* 32:139–152.
- Crowder RG, Reznick JS, Rosenkrantz SL (1991) Perception of the major/minor distinction: V. Preferences among infants. *Bull Psychon Soc* 29:187–188.
- Cuddy LL, Badertscher B (1987) Recovery of the tonal hierarchy: some comparisons across age and levels of musical experience. *Percept Psychophys* 41:609–620.
- Cunningham JG, Sterling RS (1988) Developmental change in the understanding of affective meaning in music. *Motiv Emot* 12:399–413.
- Dalla Bella S, Peretz I, Rousseau L, Gosselin N (2001) A developmental study of the affective value of tempo and mode in music. *Cognition* 80: B1–B10.
- Demany L, McKenzie B, Vurpillot E (1977) Rhythm perception in early infancy. *Nature* 266:718–719.
- Dissanayake E (2000) Antecedents of the temporal arts in early mother-infant interaction. In: Wallin NL, Merker B, Brown S (eds), *The Origins of Music*. Cambridge, MA: The MIT Press, pp. 389–410.

- Dolgin KG, Adelson EH (1990) Age changes in the ability to interpret affect in sung and instrumentally-presented melodies. *Psychol Music* 18:87–98.
- Douglas S, Willatts P (1994) The relationship between musical ability and literacy skills. *J Res Read* 17:99–107.
- Drake C (1993) Reproduction of musical rhythms by children, adult musicians, and adult nonmusicians. *Percept Psychophys* 53:25–33.
- Drake C, Penel A, Bigand E (2000a) Tapping in time with mechanically and expressively performed music. *Music Percept* 18:1–23.
- Drake C, Jones MR, Baruch C (2000b) The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. *Cognition* 77:251–288.
- Edmonston WE Jr (1966) The use of the semantic differential technique in the esthetic evaluation of musical excerpts. *Am J Psychol* 79:650–652.
- Eerola T, Luck G, Toiviainen P (August 2006) An investigation of pre-schooler' corporeal synchronization with music. Presented at the 9th International Conference on Music Perception and Cognition, Bologna, Italy.
- Fernald A (1991) Prosody in speech to children: prelinguistic and linguistic functions. *Ann Child Dev* 8:43–80.
- Fernald A (1993) Approval and disapproval: infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Dev* 64:657–674.
- Forgeard M, Schlaug G, Norton A, Rosam C, Lyengar U, Winner E (2008) The relation between music and phonological processing in normal-reading children and children with dyslexia. *Music Percept* 25:383–390.
- Foxton JM, Nandy RK, Griffiths TD (2006) Rhythm deficits in “tone deafness.” *Brain Cogn* 62:24–29.
- Fraisse P (1982) Rhythm and tempo. In: Deutsch D (ed), *The Psychology of Music*. New York: Academic Press, pp. 149–180.
- Franklin MS, Moore KS, Yip C, Jonides J, Rattray K, Moher J (2008) The effects of musical training on verbal memory. *Psychol Music* 36:353–365.
- Fujioka T, Trainor LJ, Ross B, Kakigi R, Pantev C (2004) Musical training enhances automatic encoding of melodic contour and interval structure. *J Cogn Neurosci* 16:1010–1021.
- Fujioka T, Trainor LJ, Ross B, Kakigi R, Pantev C (2005) Automatic encoding of polyphonic melodies in musicians and nonmusicians. *J Cogn Neurosci* 17:1578–1592.
- Fujioka T, Ross B, Kakigi R, Pantev C, Trainor LJ (2006) One year of musical training affects development of auditory cortical-evoked fields in young children. *Brain* 129:2593–2608.
- Fujioka T, Trainor LJ, Large EW, Ross B (2009) Beta and gamma rhythms in human auditory cortex during musical beat processing. *Ann N Y Acad Sci* 1169:89–92.
- Gabrielsson A, Juslin PN (2003) Emotional expression in music. In: Goldsmith HH, Davidson RJ, Scherer KR (eds), *Handbook of Affective Sciences*. New York: Oxford University Press, pp. 503–534.
- Gagnon L, Peretz I (2003) Mode and tempo relative contributions to “happy-sad” judgments in equitone melodies. *Cogn Emot* 17:25–40.
- Gardiner MF, Fox A, Knowles F, Jeffrey D (1996) Learning improved by arts training. *Nature* 381:284.
- Gaser C, Schlaug G (2003) Brain structures differ between musicians and non-musicians. *J Neurosci* 23:9240–9245.
- Gerardi GM, Gerken L (1995) The development of affective responses to modality and melodic contour. *Music Percept* 12:279–290.
- Gerry D, Faux A, Trainor LJ (2009) Effects of Kindermusik training on infants' perception of rhythm. *Dev Sci*. DOI: 10.1111/j.1467-7687.2009.00912.x.
- Gouzouasis P, Guhn M, Kishor N (2007) The predictive relationship between achievement and participation in music and achievement in core grade 12 academic subjects. *Music Educ Res* 9:81–92.
- Gregory AH, Worrall L, Sarge A (1996) The development of emotional responses to music in young children. *Motiv Emot* 20:341–348.

- Gromko JE (2005) The effect of music instruction on phonemic awareness in beginning readers. *J Res Music Educ* 53:199–209.
- Hannon EE, Johnson SP (2005) Infants use meter to categorize rhythms and melodies: Implications for musical structure learning. *Cogn Psychol* 50:354–377.
- Hannon EE, Trainor LJ (2007) Music acquisition: effects of enculturation and formal training on development. *Trends Cogn Sci* 11:466–472.
- Hannon EE, Trehub SE (2005a) Metrical categories in infancy and adulthood. *Psychol Sci* 16:48–55.
- Hannon EE, Trehub SE (2005b) Tuning in to musical rhythms: infants learn more readily than adults. *Proc Natl Acad Sci USA* 102:12639–12643.
- Hannon EE, Snyder JS, Eerola T, Krumhansl CL (2004) The role of melodic and temporal cues in perceiving musical meter. *J Exp Psychol Hum Percept Perform* 30:956–974.
- He C, Trainor LJ (2009) Finding the pitch of the missing fundamental in infants. *J Neurosci* 29:7718–7722.
- He C, Hotson L, Trainor LJ (2007) Mismatch responses to pitch changes in early infancy. *J Cogn Neurosci* 19:878–892.
- He C, Hotson L, Trainor LJ (2009) Maturation of cortical mismatch responses to occasional pitch change in early infancy: effects of presentation rate and magnitude of change. *Neuropsychologia* 47:218–229.
- Hepper PG, Shahidullah S (1994) The beginnings of mind: evidence from the behaviour of the fetus. *J Reprod Infant Psychol* 12:143–154.
- Hetland L (2000) Learning to make music enhances spatial reasoning. *J Aesthetic Educ* 34:179–238.
- Hevner K (1935) The affective character of the major and minor modes in music. *Am J Psychol* 47:103–118.
- Hevner K (1936) Experimental studies of the elements of expression in music. *Am J Psychol* 48:246–268.
- Ho Y, Cheung M, Chan AS (2003) Music training improves verbal but not visual memory: cross-sectional and longitudinal explorations in children. *Neuropsychology* 17:439–450.
- Huron D (2006) *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: The MIT Press.
- Husain G, Thompson WF, Schellenberg EG (2002) Effects of musical tempo and mode on arousal, mood, and spatial abilities. *Music Percept* 20:151–171.
- Hutchinson S, Lee LH, Gaab N, Schlaug G (2003) Cerebellar volume of musicians. *Cereb Cortex* 13:943–949.
- Jakobson LS, Cuddy LL, Kilgour AR (2003) Time tagging: a key to musicians' superior memory. *Music Percept* 20:307–313.
- Jakobson LS, Lewycky ST, Kilgour AR, Stoesz BM (2008) Memory for verbal and visual material in highly trained musicians. *Music Percept* 26:41–55.
- Jentschke S, Koelsch S, Friederici AD (2005) Investigating the relationship of music and language in children: influences of musical training and language impairment. *Ann NY Acad Sci* 1060:231–242.
- Jones MR, Boltz M (1989) Dynamic attending and responses to time. *Psychol Rev* 96:459–491.
- Jones MR, Yee W (1997) Sensitivity to time change: the role of context and skill. *J Exp Psychol Hum Percept Perform* 23:693–709.
- Jones MR, Jagacinski RJ, Yee W, Floyd RL, Klapp ST (1995) Tests of attentional flexibility in listening to polyrhythmic patterns. *J Exp Psychol Hum Percept Perform* 21:293–307.
- Jongsma MLA, Desain P, Honing H (2004) Rhythmic context influences the auditory evoked potentials of musicians and nonmusicians. *Biol Psychol* 66:129–152.
- Juslin PN, Laukka P (2003) Communication of emotions in vocal expression and music performance: different channels, same code? *Psychol Bull* 129:770–814.
- Juslin PN, Laukka P (2004) Expression, perception, and induction of musical emotions: a review and a questionnaire study of everyday listening. *J New Music Res* 33:217–238.

- Kastner MP, Crowder RG (1990) Perception of the major/minor distinction: IV. Emotional connotations in young children. *Music Percept* 8:189–201.
- Kivy P (1980) *The Corded Shell: Reflections on Musical Expression*. Princeton: Princeton University Press.
- Koelsch S, Siebel WA (2005) Towards a neural basis of music perception. *Trends Cogn Sci* 9:578–584.
- Koelsch S, Gunter T, Friederici AD, Schröger E (2000) Brain indices of music processing: “nonmusicians” are musical. *J Cogn Neurosci* 12:520–541.
- Koelsch S, Schmidt B, Kansok J (2002) Effects of musical expertise on the early right anterior negativity: an event-related brain potential study. *Psychophysiology* 39:657–663.
- Koelsch S, Grossmann T, Gunter TC, Hahne A, Schröger E, Friederici AD (2003) Children processing music: electric brain responses reveal musical competence and gender differences. *J Cogn Neurosci* 15:683–693.
- Kratus J (1993) A developmental study of children’s interpretation of emotion in music. *Psychol Music* 21:3–19.
- Krumhansl CL (1997) An exploratory study of musical emotions and psychophysiology. *Can J Exp Psychol* 51:336–353.
- Krumhansl CL, Keil FC (1982) Acquisition of the hierarchy of tonal functions in music. *Mem Cognit* 10:243–251.
- Kuriki S, Kanda S, Hirata Y (2006) Effects of musical experience on different components of MEG responses elicited by sequential piano-tones and chords. *J Neurosci* 26:4046–4053.
- Lappe C, Herholz SC, Trainor LJ, Pantev C (2008) Cortical plasticity induced by short-term unimodal and multimodal musical training. *J Neurosci* 28:9632–9639.
- Large EW, Jones MR (1999) The dynamics of attending: how people track time-varying events. *Psychol Rev* 106:119–159.
- Lecanuet JP, Graniere-Deferre C, Jacquet A-Y, DeCasper AJ (2000) Fetal discrimination of low-pitched musical notes. *Dev Psychobiol* 36:29–39.
- Lee Y, Lu M, Ko H (2007) Effects of skill training on working memory capacity. *Learn Instr* 17:336–344.
- Leppänen PHT, Eklund KM, Lyytinen H (1997) Event-related brain potentials to change in rapidly presented acoustic stimuli in newborns. *Dev Neuropsychol* 13:175–204.
- Leppänen PHT, Guttorm TK, Pihko E, Takkinen S, Eklund KM, Lyytinen H (2004) Maturation effects on newborn ERPs measured in the mismatch negativity paradigm. *Exp Neurol* 190:91–101.
- Lerdahl F, Jackendoff RS (1983) *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.
- Limb CJ, Kemeny S, Ortigoza EB, Rouhani S, Braun AR (2006) Left hemispheric lateralization of brain activity during passive rhythm perception in musicians. *Anat Rec A Discov Mol Cell Evol Biol* 288A:382–389.
- Luck SJ (2005) *Event-Related Potential Technique*. Cambridge: MIT Press.
- Lynch MP, Eilers RE, Oller DK, Urbano RC (1990) Innateness, experience, and music perception. *Psychol Sci* 1:272–276.
- Magne C, Schön D, Besson M (2006) Musician children detect pitch violations in both music and language better than nonmusician children: Behavioral and electrophysiological approaches. *J Cogn Neurosci* 18:199–211.
- Marques C, Moreno S, Castro SL, Besson M (2007) Musicians detect pitch violation in a foreign language better than nonmusicians: behavioral and electrophysiological evidence. *J Cogn Neurosci* 19:1453–1463.
- Masataka N (2006) Preference for consonance over dissonance by hearing newborns of deaf parents and of hearing parents. *Dev Sci* 9:46–50.
- McAuley JD, Jones MR, Holub S, Johnston HM, Miller NS (2006) The time of our lives: life span development of timing and event tracking. *J Exp Psychol Gen* 135:348–367.
- Meyer LB (1956) *Emotion and Meaning in Music*. Chicago: University of Chicago Press.

- Moreno S, Besson M (2006) Musical training and language-related brain electrical activity in children. *Psychophysiology* 43:287–291.
- Moreno S, Marques C, Santos A, Santos M, Castro SL, Besson M (2008) Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. *Cereb Cortex* 19:712–723.
- Morronegello BA (1984) Auditory temporal pattern perception in 6- and 12-month-old infants. *Dev Psychol* 20:441–448.
- Morton JB, Trehub SE (2007) Children's judgements of emotion in song. *Psychol Music* 35:629–639.
- Näätänen R, Paavilainen P, Rinne T, Alho K (2007) The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin Neurophysiol* 118:2544–2590.
- Nakata T, Trehub SE (2004) Infants' responsiveness to maternal speech and singing. *Infant Behav Dev* 27:455–464.
- Nawrot ES (2003) The perception of emotional expression in music: evidence from infants, children and adults. *Psychol Music* 31:75–92.
- Nyklíček I, Thayer JF, Van Doornen LJP (1997) Cardiorespiratory differentiation of musically-induced emotions. *J Psychophysiol* 11:304–321.
- Overy K (2003) Dyslexia and music: from timing deficits to musical intervention. *Ann NY Acad Sci* 999:497–505.
- Pantev C, Oostenveld R, Engelien A, Ross B, Roberts LE, Hoke M (1998) Increased auditory cortical representation in musicians. *Nature* 392:811–814.
- Pantev C, Roberts LE, Schulz M, Engelien A, Ross B (2001) Timbre-specific enhancement of auditory cortical representations in musicians. *NeuroReport* 12:169–174.
- Papoušek M (1992) Early ontogeny of vocal communication in parent-infant interactions. In: Papoušek H, Jürgens U, Papoušek M (eds), *Nonverbal Vocal Communication: Comparative and Developmental Approaches*. New York: Cambridge University Press, pp. 230–261.
- Patel AD (2008) *Music, Language, and the Brain*. Oxford: Oxford University Press.
- Peretz I (2001) Listen to the brain: A biological perspective on musical emotions. In: Juslin PN, Sloboda JA (eds), *Music and Emotion: Theory and Research*. New York: Oxford University Press, pp. 105–134.
- Phillips-Silver J, Trainor LJ (2005) Feeling the beat: movement influences infant rhythm perception. *Science* 308:1430.
- Phillips-Silver J, Trainor LJ (2007) Hearing what the body feels: auditory encoding of rhythmic movement. *Cognition* 105:533–546.
- Picton TW, Alain C, Otten L, Ritter W, Achim A (2000) Mismatch negativity: different water in the same river. *Audiol Neurootol* 5:111–139.
- Plantinga J, Trainor LJ (2005) Memory for melody: infants use a relative pitch code. *Cognition* 98:1–11.
- Plantinga J, Trainor LJ (2008) Infants' memory for isolated tones and the effects of interference. *Music Percept* 26:121–127.
- Plantinga J, Trainor LJ (2009) Melody recognition by two-month-old infants. *J Acoust Soc Am* 125:EL58–EL62.
- Ponton CW, Eggermont JJ, Kwong B, Don M (2000) Maturation of human central auditory system activity: evidence from multi-channel evoked potentials. *Clin Neurophysiol* 111:220–236.
- Rauscher FH, Shaw GL, Ky CN (1993) Music and spatial task-performance. *Nature* 365:611.
- Repp BH (2005) Sensorimotor synchronization: a review of the tapping literature. *Psychon Bull Rev* 12:969–992.
- Repp BH, London J, Keller PE (2005) Production and synchronization of uneven rhythms at fast tempi. *Music Percept* 23:61–78.
- Robazza C, Macaluso C, D'Urso V (1994) Emotional reactions to music by gender, age, and expertise. *Percept Motor Skills* 79:939–944.
- Rock AML, Trainor LJ, Addison TL (1999) Distinctive messages in infant-directed lullabies and play songs. *Dev Psychol* 35:527–534.

- Romand R (1992) *Development of Auditory and Vestibular Systems*. New York: Elsevier.
- Ross DA, Olson IR, Marks LE, Gore JC (2004) A nonmusical paradigm for identifying absolute pitch possessors. *J Acoust Soc Am* 116:1793–1799.
- Saffran JR, Griepentrog GJ (2001) Absolute pitch in infant auditory learning: evidence for developmental reorganization. *Dev Psychol* 37:74–85.
- Schachner A, Brady TF, Pepperberg IM, Hauser M (June 2008) Spontaneous entrainment to auditory rhythms in vocal-learning bird species. Presented at The Neurosciences and Music III, Montreal, Canada.
- Schellenberg EG (2001) Music and nonmusical abilities. *Ann NY Acad Sci* 930:355–371.
- Schellenberg EG (2004) Music lessons enhance IQ. *Psychol Sci* 15:511–514.
- Schellenberg EG (2005) Music and cognitive abilities. *Curr Dir Psychol Sci* 14:317–320.
- Schellenberg EG (2006) Long-term positive associations between music lessons and IQ. *J Educ Psychol* 98:457–468.
- Schellenberg EG, Trainor LJ (1996) Sensory consonance and the perceptual similarity of complex-tone harmonic intervals: tests of adult and infant listeners. *J Acoust Soc Am* 100:3321–3328.
- Schellenberg EG, Bigand E, Poulin-Charronnat B, Garnier C, Stevens C (2005) Children's implicit knowledge of harmony in western music. *Dev Sci* 8:551–566.
- Scherer KR (1985) Vocal affect signaling: a comparative approach. *Adv Study Behav* 15:189–244.
- Scherer KR (1986) Vocal affect expression: a review and a model for future research. *Psychol Bull* 99:143–165.
- Schlaug G (2009) Music, musicians, and brain plasticity. In: Hallam S, Cross I, Thaut M (eds), *The Oxford Handbook of Music Psychology*. Oxford: Oxford University Press, pp. 197–207.
- Schmidt LA, Trainor LJ (2001) Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions. *Cogn Emot* 15:487–500.
- Schmidt LA, Trainor LJ, Santesso DL (2003) Development of frontal electroencephalogram (EEG) and heart rate (ECG) responses to affective musical stimuli during the first 12 months of post-natal life. *Brain Cogn* 52:27–32.
- Schneider P, Scherg M, Dosch HG, Specht HJ, Gutschalk A, Rupp A (2002) Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nat Neurosci* 5:688–694.
- Schön D, Magne C, Besson M (2004) The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology* 41:341–349.
- Scott L (1992) Attention and perseverance behaviors of preschool children enrolled in Suzuki violin lessons and other activities. *J Res Music Educ* 40:225–235.
- Shahin A, Roberts LE, Trainor LJ (2004) Enhancement of auditory cortical development by musical experience in children. *NeuroReport* 15:1917–1921.
- Shahin A, Roberts LE, Pantev C, Trainor LJ, Ross B (2005) Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. *NeuroReport* 16:1781–1785.
- Shahin AJ, Roberts LE, Chau W, Trainor LJ, Miller LM (2008) Music training leads to the development of timbre-specific gamma band activity. *Neuroimage* 41:113–122.
- Sloboda JA (1985) *The Musical Mind: The Cognitive Psychology of Music*. Oxford: Clarendon Press.
- Sloboda JA (1991) Music structure and emotional response: some empirical findings. *Psychol Music* 19:110–120.
- Sluming V, Barrick T, Howard M, Cezayirli E, Mayes A, Roberts N (2002) Voxel-based morphometry reveals increased gray matter density in Broca's area in male symphony orchestra musicians. *Neuroimage* 17:1613–1622.
- Snyder J, Krumhansl CL (2001) Tapping to ragtime: Cues to pulse finding. *Music Percept* 18:455–489.
- Snyder JS, Large EW (2005) Gamma-band activity reflects the metric structure of rhythmic tone sequences. *Cogn Brain Res* 24:117–126.
- Snyder JS, Hannon EE, Large EW, Christiansen MH (2006) Synchronization and continuation tapping to complex meters. *Music Percept* 24:135–145.
- Speer JR, Meeks PU (1985) School children's perception of pitch in music. *Psychomusicology* 5:49–56.

- Terwogt MM, van Grinsven F (1991) Musical expression of moodstates. *Psychol Music* 19:99–109.
- Tew S, Fujioka T, Trainor LJ (2009) Neural representation of transposed melody at 6 months of age. *Ann N Y Acad Sci* 1169:287–290.
- Thompson WF, Schellenberg EG, Husain G (2004) Decoding speech prosody: do music lessons help? *Emotion* 4:46–64.
- Tillmann B, Bigand E, Madurell F (1998) Local versus global processing of harmonic cadences in the solution of musical puzzles. *Psychol Res* 61:157–174.
- Tillmann B, Bigand E, Escoffier N, Lalitte P (2006) The influence of musical relatedness on timbre discrimination. *Eur J Cogn Psychol* 18:343–358.
- Todd NP, Cousins R, Lee CS (2007) The contribution of anthropometric factors to individual differences in the perception of rhythm. *Emp Musicol Rev* 2:1–13.
- Trainor LJ (1996) Infant preferences for infant-directed versus noninfant-directed playsongs and lullabies. *Infant Behav Dev* 19:83–92.
- Trainor LJ (1997) Effect of frequency ratio on infants' and adults' discrimination of simultaneous intervals. *J Exp Psychol Hum Percept Perform* 23:1427–1438.
- Trainor LJ (2005) Are there critical periods for musical development? *Dev Psychobiol* 46:262–278.
- Trainor LJ (2007) Do preferred beat rate and entrainment to the beat have a common origin in movement? *Emp Musicol Rev* 2:17–20.
- Trainor LJ (2008) Science and music: the neural roots of music. *Nature* 453:598–599.
- Trainor LJ, Adams B (2000) Infants' and adults' use of duration and intensity cues in the segmentation of tone patterns. *Percept Psychophys* 62:333–340.
- Trainor LJ, Heinmiller BM (1998) The development of evaluative responses to music: infants prefer to listen to consonance over dissonance. *Infant Behav Dev* 21:77–88.
- Trainor LJ, Schmidt LA (2003) Processing emotions induced by music. In: Peretz I, Zatorre R (eds), *The Cognitive Neuroscience of Music*. New York: Oxford University Press, pp. 311–324.
- Trainor LJ, Trehub SE (1992a) A comparison of infants' and adults' sensitivity to Western musical structure. *J Exp Psychol Hum Percept Perform* 18:394–402.
- Trainor LJ, Trehub SE (1992b) The development of referential meaning in music. *Music Percept* 9:455–470.
- Trainor LJ, Trehub SE (1993) What mediates infants' and adults' superior processing of the major over the augmented triad? *Music Percept* 11:185–196.
- Trainor LJ, Trehub SE (1994) Key membership and implied harmony in Western tonal music: Developmental perspectives. *Percept Psychophys* 56:125–132.
- Trainor LJ, Zatorre R (2009) The neurobiological basis of musical expectations: from probabilities to emotional meaning. In: Hallam S, Cross I, Thaut M (eds), *Oxford Handbook of Music Psychology*. Oxford University Press, pp. 171–183.
- Trainor LJ, Clark ED, Huntley A, Adams BA (1997) The acoustic basis of preferences for infant-directed singing. *Infant Behav Dev* 20:383–396.
- Trainor LJ, Desjardins RN, Rockel C (1999) A comparison of contour and interval processing in musicians and nonmusicians using event-related potentials. *Aust J Psychol* 51:147–153.
- Trainor LJ, Austin CM, Desjardins RN (2000) Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychol Sci* 11:188–195.
- Trainor LJ, Tsang CD, Cheung VHW (2002a) Preference for sensory consonance in 2- and 4-month-old infants. *Music Percept* 20:187–194.
- Trainor LJ, McDonald KL, Alain C (2002b) Automatic and controlled processing of melodic contour and interval information measured by electrical brain activity. *J Cogn Neurosci* 14:430–442.
- Trainor LJ, Shahin AJ, Roberts LE (2003) Effects of musical training on the auditory cortex in children. *Ann NY Acad Sci* 999:506–513.
- Trainor LJ, Gao X, Lei J, Lehtovarara K, Harris LR (2009) The primal role of the vestibular system in determining musical rhythm. *Cortex* 45:35–43.

- Tramo MJ, Cariani PA, Delgutte B, Braida LD (2001) Neurobiological foundations for the theory of harmony in Western tonal music. *Ann NY Acad Sci* 930:92–116.
- Trehub SE (2000) Human processing predispositions and musical universals. In: Wallin NL, Merker B, Brown S (eds), *The Origins of Music*. Cambridge, MA: The MIT Press, pp. 427–448.
- Trehub SE (2003) Musical predispositions in infancy: an update. In: Peretz I, Zatorre R (eds), *The Cognitive Neuroscience of Music*. New York: Oxford University Press, pp. 3–20.
- Trehub SE, Thorpe LA (1989) Infants' perception of rhythm: categorization of auditory sequences by temporal structure. *Can J Psychol* 43:217–229.
- Trehub SE, Trainor LJ (1993) Listening strategies in infancy: the roots of music and language development. In: McAdams S, Bigand E (eds), *Thinking in Sound: The Cognitive Psychology of Human Audition*. New York: Oxford University Press, pp. 278–327.
- Trehub SE, Trainor LJ (1998) Singing to infants: lullabies and play songs. *Advances in Infancy Research* 12:43–77.
- Trehub SE, Bull D, Thorpe LA (1984) Infants' perception of melodies: the role of melodic contour. *Child Dev* 55:821–830.
- Trehub SE, Cohen AJ, Thorpe LA, Morrongiello BA (1986) Development of the perception of musical relations: semitone and diatonic structure. *J Exp Psychol Hum Percept Perform* 12:295–301.
- Trehub SE, Unyk AM, Trainor LJ (1993a) Adults identify infant-directed music across cultures. *Infant Behav Dev* 16:193–211.
- Trehub SE, Unyk AM, Trainor LJ (1993b) Maternal singing in cross-cultural perspective. *Infant Behav Dev* 16:285–295.
- Trehub SE, Schellenberg EG, Kamenetsky SB (1999) Infants' and adults' perception of scale structure. *J Exp Psychol Hum Percept Perform* 25:965–975.
- Vaughn K (2000) Music and mathematics: modest support for the oft-claimed relationship. *J Aesthetic Educ* 34:149–166.
- Volkova A, Trehub SE, Schellenberg EG (2006) Infants' memory for musical performances. *Dev Sci* 9:583–589.
- Vuust P, Pallesen KJ, Bailey C, van Zuijen TL, Gjedde A, Roepstorff A, et al. (2005) To musicians, the message is in the meter: pre-attentive neuronal responses to incongruent rhythm are left-lateralized in musicians. *Neuroimage* 24:560–564.
- Werker JF, McLeod PJ (1989) Infant preference for both male and female infant-directed talk: a developmental study of attentional and affective responsiveness. *Can J Psychol* 43:230–246.
- Werner LA, Marean GC (1996) *Human Auditory Development*. Madison, WI: Brown Benchmark.
- Winkler I, Kushnerenko E, Horváth J, Čeponienė R, Fellman V, Huotilainen M, et al. (2003) Newborn infants can organize the auditory world. *Proc Natl Acad Sci USA* 100:11812–11815.
- Winkler I, Háden GP, Ladinig O, Sziller I, Honing H (2009) Newborn infants detect the beat in music. *Proc Natl Acad Sci USA*, 106:2468–2471.
- Zatorre RJ (2001) Neural specializations for tonal processing. *Ann NY Acad Sci* 930:193–210.
- Zentner MR, Kagan J (1998) Infants' perception of consonance and dissonance in music. *Infant Behav Dev* 21:483–492.

Chapter 5

Music and Emotion

Patrick G. Hunter and E. Glenn Schellenberg

5.1 Introduction to the Study of Music and Emotion

Music is the shorthand of emotion.

–Leo Tolstoy

Why waste money on psychotherapy when you can listen to the B Minor Mass?

–Michael Torke

These two quotations reflect common attitudes about music. Tolstoy's comment suggests that music *conveys* emotion, whereas Torke's question implies that music *influences* listeners' emotions. Section 5.2 of the present chapter includes a discussion of the various theoretical approaches that are used to explain affective responses to music. Few scholars dispute the claim that listeners recognize emotions in music. Some argue, however, that music does not elicit true emotions in the listener (e.g., Kivy 1980, 1990, 2001). For example, many years ago Meyer (1956) posited that affective responses to music consist of experiences of tension and relaxation (rather than actual emotions), which occur when listeners' expectancies about what will happen next in a piece of music are violated or fulfilled, respectively. This position has been challenged in recent years with findings from studies using behavioral, physiological, and neurological measures, all of which indicate that listeners respond affectively to music (e.g., Krumhansl 1997; Gagnon and Peretz 2003; Mitterschiffthaler et al. 2007; Witvliet and Vrana 2007). Nonetheless, the debate continues (e.g., Konečni 2008).

Even if one accepts that listeners respond emotionally to music, there are still disagreements about the nature of the response. Section 5.2 also includes an examination of issues that remain unresolved about whether emotional responses to music are

E.G. Schellenberg (✉)

Department of Psychology, University of Toronto at Mississauga, Mississauga,
ON, Canada L5L 1C6

e-mail: g.schellenberg@utoronto.ca

(1) “real” emotions, (2) a separate class of “aesthetic” emotions, or (3) moods rather than emotions. A discussion of how to characterize and classify emotional responses to music is included. Categorical (or discrete) models of emotions, drawing from research on *basic emotions* (Ekman 1984, 1992), have been used in much of the past research. By contrast, dimensional (or continuous) models of emotion describe emotions in terms of various dimensions. Although the two approaches are not completely incompatible, the fit is not perfect. For example, the *circumplex model* (Russell 1980) uses two dimensions to describe emotions: *valence* (ranging from displeasure to pleasure) and *arousal* (level of activation). This approach works well for some specific emotions (e.g., *joy* has positive valence and high arousal; *sadness* has negative valence and low arousal), but others are undifferentiated on these dimensions. For example, *fear* and *anger* both have negative valence and high arousal.

Are certain musical characteristics such as tempo, mode, and loudness reliable indicators of a piece’s associations with one or more emotions? Are some characteristics more consistent than others in being associated with emotions across listeners and musical styles? Are some emotions easier than other emotions to convey and induce with music? Do musical characteristics interact in their influence on the emotional status of a piece? Section 5.3 provides a review of evidence relevant to these sorts of questions.

Various ways to measure emotional responses to music are described and critiqued in Sect. 5.4. There are as many approaches to measurement of emotions as there are theories about what these are. Different models of emotion lend themselves to different measurement techniques. Categorical models make the use of distinct labels appropriate (e.g., happy, sad, etc.), whereas dimensional models are consistent with the use of rating scales (often for arousal and valence). Both of these methods rely primarily on self-report, which may be susceptible to response bias. Technological advances have provided more “objective” ways of measuring emotional responses (e.g., physiological responses, brain imaging), but these methods are relatively insensitive to emotional responses other than arousal and pleasantness/unpleasantness (or liking/disliking).

How are listeners’ emotional responses to music similar to and different from their perceptions of the emotions conveyed by music? Differences in methods across studies make it difficult to arrive at general conclusions, or to resolve apparent discrepancies in findings that have been reported. Often, listeners are asked only about their own emotional experience (e.g., Hunter et al. 2008a), or what emotion they hear in the music (e.g., Gagnon and Peretz 2003). At other times, the locus of the emotion (i.e., felt or perceived) may be ambiguous. While emotional reactions are likely to parallel perceptions of emotions in many instances, at other times the two types of response may diverge. Section 5.5 provides a review of studies that explored links between perceiving and feeling emotions in musical contexts.

Section 5.6 addresses a specific, relatively intense emotional response to music: *chills*. First described by Goldstein (1980), chills are the “tingly” feelings that listeners sometimes experience. Chills are usually pleasurable experiences that can be accompanied by physiological reactions, such as piloerection (goosebumps).

Although the experiences are clearly emotional, it is unclear how they are evoked by music and how they fit within models of emotion. Section 5.6 explores what is known about the experience of chills (e.g., phenomenological, physiological, and neurological reactions) and reviews studies linking musical features to such experiences.

Section 5.7 comprises a discussion of one of the most basic affective responses to music: *liking*. Liking for music is reviewed as a function of consonance and dissonance, familiarity with the piece, and the emotion portrayed by the piece. Finally, Sect. 5.8 provides an overview of the chapter and highlights interesting questions that could be examined in future research.

5.2 What Are Musical Emotions?

Although scholars agree that music can *sound* happy or sad, there is contention over whether music truly evokes emotions. Even those who agree that music evokes emotions often disagree over the nature of those emotions and how they are induced. Are musically induced emotions the same as everyday emotions such as happiness, sadness, anger, and so on? Much of this debate centers on the definition of *emotion*. Outside the music literature, probably the most common position comes from *appraisal theory* (see Smith et al. 1993), which asserts that emotions result from cognitive appraisals of a target. For example, sadness is elicited by news of the death of a friend, which is appraised as beyond one's control and contrary to one's desires. The debate also depends on what "emotions" are considered to comprise. Most agree that emotions consist of a subjective feeling, but some expand the definition to include a combination of additional components, such as cognitive appraisal, physiological arousal, motor expression, and behavioral tendency (Scherer 2004). The present section details the various positions related to the presence and nature of emotional responses to music.

5.2.1 *Emotivist and Cognitivist Positions*

Music philosophers were among the first to debate the existence of music-induced emotions, with believers and nonbelievers referred to as *emotivists* and *cognitivists*, respectively. Kivy (1980, 1990, 2001), one of the main proponents of the cognitivist position, argues that happy- and sad-sounding musical pieces do not evoke true happiness and sadness in listeners. Rather, affective responses stem from listeners' evaluations of the music. He writes, "I experience unalloyed joy when I listen to sad music that is great music, utter boredom when it is sad music that is bad music" (Kivy 2001, p. 147). In line with appraisal theorists (e.g., Smith et al. 1993), Kivy argues that emotions require a cognitive appraisal of a target and that there is no target in music except the music itself. As a result, listeners simply feel positive or

negative when they like or dislike the music, respectively. In other words, listeners refer to music as *happy* or *sad* because the music expresses happiness or sadness, not because the music makes them feel happy or sad.

By contrast, most emotivist theories suggest that music actually evokes or induces feelings in listeners (for a review, see Davies 2001). Various attempts have been made to deal with the difficulty of explaining emotional reactions to music in terms of cognitive appraisals. Some scholars deny that emotions necessarily involve appraisals (e.g., Maddel 2002), and argue that other mechanisms can give rise to musical emotions. For example, emotional responses to music may be a sympathetic response (Ridley 1995; Levinson 1996). A piece of music may invoke a hypothetical person expressing emotion; the listener consequently feels a similar emotion. In a related vein, Davies (2001) agrees that listeners' feelings may mirror those expressed by the music even though they are not targeted at the music. Rather, they are experienced *contagiously*, much like being around a sad person or a group of sad people can lead to feelings of sadness. Juslin and Västfjäll (2008) argue that cognitive appraisals are but one way emotions are induced, and they propose six other mechanisms that explain how musical pieces (and other stimuli) induce emotion: (1) brain stem reflexes (e.g., reactions to dissonance), (2) conditioning (i.e., a particular piece or genre is associated with a positive or negative emotion), (3) contagion (i.e., perceptions spread to feelings, as noted), (4) visual imagery (i.e., images evoked by music act as cues to an emotion), (5) episodic memory (i.e., a piece is associated with a particular event, which, in turn, is associated with an emotion), and (6) expectancies that are fulfilled or denied (from Meyer 1956).

Huron's (2006) ITPRA (Imagination-Tension-Prediction-Response-Appraisal) theory expands upon Meyer's (1956) work on the affective consequences of expectancies. Although the theory is formulated as a general theory of expectancy, Huron applies it specifically to music. He identifies five expectancy responses, two occurring before the onset of the event and three afterwards. The first is the *imagination response*, which is somewhat removed from the event and consists of the prediction of what will happen – and how the listener will feel – when and after the musical event occurs. By contrast, the *tension response* refers to listeners' mental and physiological preparation when the expected event is imminent. After the event has occurred, listeners receive some pleasure or displeasure from the accuracy of their prediction, which is the *prediction response*. Listeners also evaluate the pleasantness or unpleasantness of the outcome, which gives rise to *reaction response*. Thus, immediately following an event that was negative but nevertheless predicted, the listener may feel some mix of pleasure and displeasure. Finally, the *appraisal response* arises with the activation of conscious thought and involves a higher-level evaluation of the event and its consequences. The entire process can lead to specific affective responses. When expectancies are met, music listeners get a certain degree of pleasure, which is reinforced if the event is positive. Nonetheless, expectancies that are unfulfilled are not necessarily negative. If the event is appraised as positive overall, the result might be laughter, awe, or chills. These responses are related to the flight, freeze, and fight responses, which occur in response to negatively appraised events.

In a different view, Matravers (1998) accepts that emotions necessarily involve cognitive appraisals. Accordingly, he proposes that the affective experience of music is not one of emotions per se. Rather, musical affect consists solely of a feeling component – the subjectively felt component of an emotion without the associated cognition. This proposal blurs the distinction between emotions and *moods*, which are typically used to describe diffuse but relatively long-lasting feelings with no clear target (e.g., Morris 1992; Clore et al. 1994; Frijda 1994; Russell 2003; Schimmack and Crites, 2005). To further complicate matters, others argue that moods are dispositions towards certain kinds of cognitions (Frijda 1993; Siemer 2001). For example, when participants are asked to recall a sad or angry memory, the number of mood-congruent thoughts determines the degree to which the corresponding mood is experienced (Siemer 2005). When moods are measured after listening to happy-, sad-, or angry-sounding music (Siemer 2005), however, correlations between mood and thought processes could arise because the music initially evoked the mood, which then influenced participants' way of thinking. Moreover, the implication that sad music (or any music expressing negative emotions) induces sad feelings because it evokes sad thoughts is problematic. Why would listeners choose to hear sad-sounding music if it induces sad thoughts?

Regardless of whether music induces emotions or moods, there is much evidence against a strict cognitivist position. Physiological and neurological responses to music are similar to those that accompany emotional responding in general (e.g., Krumhansl 1997; Nyklíček et al. 1997; Mitterschiffthaler et al. 2007; Witvliet and Vrana 2007). In a recent study, however, physiological and self-report ratings were not completely synchronized during music listening (Grewe et al. 2007a). Because emotions are often considered to have physiological, psychological, and behavioral components (Scherer 2004), the authors interpreted the lack of synchronicity as evidence that music did not actually induce emotional responding. Nonetheless, these components are often out-of-sync in response to nonmusical events that elicit emotions (Niedenthal et al. 2006). In short, because music can elicit each component of emotion, it seems relatively safe to conclude that music induces some sort of emotional responding.

An alternative perspective (Scherer 2004; Zentner et al. 2008) argues that emotional reactions to music are common but that these differ from normal conceptions of emotion. For example, feelings of transcendence are a relatively common consequence of music listening, yet transcendence does not map readily into two-dimensional space defined by arousal and valence, and it is quite dissimilar from prototypical emotions used in the categorical approach. Thus, music is said to elicit a separate class of *aesthetic emotions* (see also Konečni 2008) that are distinct from everyday or *utilitarian emotions*. For aesthetic emotions, the feeling component is obvious but the behavioral and physiological components are often obscure. In one study that used retrospective self reports (Zentner et al. 2008), participants listed how frequently they perceive and feel a large number of affective terms in response to music. Principal components analysis revealed that common affective responses to music could be grouped into one of nine categories: *Wonder, Transcendence, Tenderness, Nostalgia, Peacefulness, Power, Joyful Activation, Tension, and Sadness*. Although there is some

overlap between these terms and the terms used in categorical and dimensional models, these aesthetic emotions also differ substantially from everyday emotions.

Perhaps the strongest evidence for affective responding to music comes from the many studies that use music to induce moods (for review, see Västfjäll 2002). Presumably, the method would not be so common if it did not work. Self-reports confirm that listeners' moods are influenced subjectively by music listening, whereas measurable effects on cognition provide more objective evidence. One example, the so-called *Mozart effect* (Rauscher et al. 1993), refers to the finding that listening to music composed by Mozart improves performance on tests of spatial abilities. Follow-up studies reveal that similar enhancements are evident following exposure to other pleasant stimuli and on tests that measure nonspatial abilities, whereas lower levels of performance are observed after exposure to less appealing stimuli (Nantais and Schellenberg 1999; Ivanov and Geake 2003; Schellenberg and Hallam 2005). More importantly, the link between exposure to music and cognitive performance is mediated by the listener's mood and arousal level (Thompson et al. 2001; Husain et al. 2002; Schellenberg et al. 2007).

When all of the available evidence is considered, it is clear that music listening often leads to emotional responses that are more complex than simple liking and disliking. Nonetheless, music-induced affective responses may differ from common definitions of emotion, both in quality and because they are not directed at the source. Moreover, it is difficult to explain why people often choose to listen to sad-sounding music. Because negative emotions are usually associated with avoidance motivation (Davidson 1998), one would expect sad-sounding music to be avoided. The bottom line is that whether affective reactions to music should be called true "emotions" is largely a question of semantics. At the very least, the evidence demonstrates that music listening influences physiological and neurological indices of emotion, and listeners report feeling these emotions.

5.2.2 *The Structure of Emotions*

Despite the apparent categorical nature of emotions, some scholars argue that all emotions can be reduced to less specific core affect (e.g., pleasure and displeasure; Ortony and Turner 1990). Others (e.g., Russell 1980, 2003) suggest that affective experience can be explained largely using two continuous dimensions: arousal (high and low) and valence (positive and negative). These dimensions of the circumplex model are thought to be orthogonal, such that any emotion can be characterized by its coordinates in a two-dimensional space (see Fig. 5.1, left panel). To illustrate, happiness usually has positive valence and moderately high arousal, whereas sadness has negative valence and moderately low arousal. Support for this two-dimensional approach comes from the International Affective Picture System, a set of widely used visual stimuli that are used to represent and evoke emotions (Lang et al. 1997). The stimulus pictures are classified according to their arousal and valence. The bipolar valence dimension also helps to explain

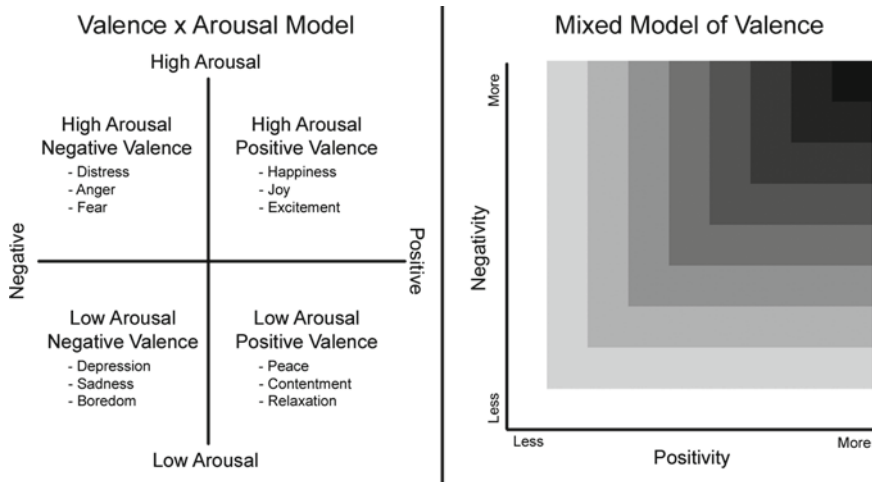


Fig. 5.1 Schematic illustrations of two models of emotion. The *left panel* shows a diagram of two-dimensional affective (valence X arousal) space – the circumplex model. Example emotions are noted in each quadrant. The *right panel* shows a model of mixed valence. Pure positive and negative responses lie along the axes in white. Darker shades of gray represent greater mixed feelings, which have shared positive and negative activation to varying degrees

affective influences on the *startle response*, the blinking reflex elicited by unexpected loud sounds (see Lang 1995). High arousal and positive valence lead to an attenuated response, whereas high arousal and negative valence lead to an exaggerated response. Fontaine et al. (2007) argue, however, that two dimensions are insufficient to categorize emotional responses completely. In a cross-cultural study, their participants were asked to differentiate 24 emotions. Responses were best described in terms of *four* rather than two dimensions: *evaluation-pleasantness*, *activation-arousal*, *potency-control*, and *unpredictability*. In short, differences in arousal and valence may fail to capture relevant distinctions among some emotions. For example, fear and stress are both negatively valenced and high on arousal but they are characterized by high and low unpredictability, respectively.

In music research, the two-dimensional model has been applied widely with considerable success (e.g., Krumhansl 1997; Schmidt and Trainor 2001; Thompson et al. 2001; Husain et al. 2002; Kreutz et al. 2008; Vieillard et al. 2008). In one study, multidimensional scaling was used to examine the underlying structure of emotional responses to music (Bigand et al. 2005). Listeners were asked to group pieces on the basis of their similarity in emotional meaning. The groupings revealed two main dimensions, arousal and valence, and a weaker third dimension related to kinetics. In an earlier study that also used multidimensional scaling (Wedin 1972), listeners rated musical excerpts on a number of different emotions. Again, the scaling solution revealed three dimensions: intensity–softness, pleasantness–unpleasantness, and solemnity–triviality. The first two dimensions corresponded closely to arousal and valence, respectively. Both of these studies provide some support for

the two-dimensional approach, but they also imply that two dimensions may not be enough to explain completely the emotions expressed by music.

Moreover, although the dimensional approach has been used extensively as a framework for musical and nonmusical stimuli, mainstream emotion researchers have noted that these models are inadequate at explaining some common emotional responses, particularly those involving ambiguity or mixed feelings. Dimensional models assume that positive and negative valence lie on opposite ends of a bipolar dimension (Russell 1980; Fontaine et al. 2007). Thus, positive and negative emotions are mutually exclusive and cannot be felt simultaneously (Russell and Carroll 1999). By contrast, the *evaluative space model* (Cacioppo and Berntson 1994) suggests that positive and negative valence can be coactivated under some circumstances (see Fig. 5.1, right panel). This alternative conceptualization has received empirical support from behavioral studies (Diener and Iran-Nejad 1986; Larsen et al. 2001, 2004, 2009; Schimmack 2001; Hunter et al. 2008a). For example, in one study, participants performed a gambling task in which they could either win or lose either a small or a large amount (Larsen et al. 2004). When participants won or lost the larger amounts, they reported unambiguous positive and negative feelings, respectively. When they won or lost the smaller amounts, however, they reported feeling both positive *and* negative affect. In the case of winning the smaller amount, they appeared to feel positive that they had won, but negative that they did not win the larger amount. Conversely, when they lost the smaller amount, they felt negative about losing but positive that they did not lose even more. Happiness and sadness – putative opposites in valence – are also linked to neural substrates that are at least partially separable (Damasio et al. 2000). If valence were a single bipolar dimension, one would predict that happiness and sadness are subserved by degree of activation in a single substrate, or by two substrates that are mutually inhibitory.

Evidence of mixed feelings has also been found in response to music (Hunter et al. 2008a), a finding that is consistent with phenomenological experience and evaluations of what makes a musical piece interesting. In two experiments, participants listened to 30-s excerpts from musical recordings that had cues to happiness (fast tempo and major mode), sadness (slow tempo and minor mode), or to both happiness and sadness (fast tempo and minor mode, or slow tempo and major mode). In one experiment, participants rated their emotional responses on two separate unipolar scales: one for happiness and one for sadness, with both scales ranging from *not at all* to *extremely*. In a second experiment, they provided a single response on a two-dimensional grid, with one axis corresponding to happiness and the other to sadness (see also Larsen et al. 2009). In both experiments, participants reported greater levels of simultaneous happy *and* sad feelings when the tempo and mode cues were mixed compared to when they were consistent. Another interesting finding was that sad-sounding music elicited higher levels of mixed feelings compared to happy-sounding music.

Response patterns similar to those reported by Hunter et al. (2008a) were evident in a separate study that used more controlled stimuli: MIDI versions of Bach pieces that were manipulated with computer software to sound happy (fast and major), sad (slow and minor), or mixed (fast and minor, or slow and major; Hunter et al. 2010). Again, participants were more likely to respond ambiguously to music with mixed

cues, and sad music elicited a larger degree of mixed feelings than happy music. Considered jointly, these studies demonstrate that music can elicit mixed feelings reliably and predictably, a finding that is precluded when happiness and sadness are measured with a single bipolar rating scale that ranges from *extremely sad* to *extremely happy*. Indeed, when listeners are asked to rate mixed feelings on a bipolar scale, they are compelled to either ignore the weaker response or to average their feelings, which leads to a relatively neutral rating that is insensitive to the positive *and* the negative aspects of their emotional state.

5.2.3 Summary

Section 5.2 was concerned with debates that extend beyond the music psychology literature. The question of which components should be included under the definition of “emotion” remains unresolved, and the choice of definition determines whether affective reactions to music can be considered “emotions.” Some scholars include the subjective feeling only, whereas others include associated cognitions, and still others include physiological changes, motor changes, and behavioral tendencies. Although all of these components can be activated in response to music, they do not necessarily covary synchronously with each other. Moreover, the majority of affective reactions to music do not appear to be mediated cognitively in the same way that they are in response to many non-musical stimuli or events. For example, listening to Beethoven’s Ninth Symphony or to The Rolling Stones’ “Jumpin’ Jack Flash” may cause strong positive feelings, yet these responses are not typically mediated by cognitive appraisals of Beethoven or The Rolling Stones, respectively. Thus, various alternatives have been proposed. Affective responses to music might consist of short-lasting moods, or they may be elicited through emotional contagion (i.e., perceived emotion in the music causes felt mood). Alternatively, music may elicit a separate class of aesthetic emotions that differ qualitatively from everyday emotions.

The second issue raised in this section concerned the use of the circumplex model as an explanatory framework for affective responses to music. Studies using nonmusical stimuli have provided evidence against the proposal that valence is a strictly bipolar dimension. Indeed, the use of bipolar ratings scales “can allow ambivalence to masquerade as neutrality by preventing respondents from reporting that they feel both good and bad” (Larsen et al. 2009, p. 454). In line with this view, studies using musical stimuli indicate that mixed feelings may be a relatively common and predictable result of music listening.

5.3 Emotions and Musical Characteristics

There have been many attempts to link emotions with specific aspects or dimensions of music, and to examine the consistency of these associations. Music varies on several dimensions (tempo, mode, loudness, pitch height, and so on) that are likely to influence emotional responding. In a classic series of studies

(Hevner 1935, 1936, 1937), listeners heard music that varied systematically along a number of these dimensions (e.g., mode, tempo, pitch). From a list of emotional adjectives, they choose the ones that best fit the music. Tempo and mode were the strongest determinants of perceived emotion in music. By asking listeners to choose emotions that “fit” the music, however, it is unclear whether their selections were based on their perception of emotion expressed by the music, or on their own feelings evoked by the music.

In other studies of musical correlates for different emotions, judgments of happiness and sadness tend to be more consistent than other emotions, such as fear and anger (Terwogt and Van Grinsven 1991; Gabrielsson and Juslin 1996; Krumhansl 1997), probably because of relatively straightforward associations with tempo and mode. Specifically, fast and slow tempos are associated with happiness and sadness, respectively, as are major and minor modes (for reviews see Gabrielsson and Juslin 2003; Juslin and Laukka 2004).

Far fewer studies have examined associations between affective responses and other musical dimensions, such as loudness, timbre, and pitch height. Gundlach (1935) reported that louder pieces were described as more animated, brilliant, uneasy, triumphant, and exalted, but less tranquil, mournful, melancholy, delicate, and sentimental. In a cross-cultural study (Balkwill et al. 2004), loudness was associated positively with perceptions of anger across Western, Japanese, and Hindustani musical styles, which suggests that loudness might be a universal cue to anger. Louder music is also predictive of higher levels of perceived activation and tension (Ilie and Thompson 2006). For female listeners, louder music may actually evoke negative feelings (Kellaris and Rice 1993). *Changes* in loudness are also important cues to emotion. Crescendos and decrescendos are associated with increases and decreases in arousal (Schubert 2004), respectively, whereas chills may be induced by crescendos (Panksepp 1995; Nagel et al. 2008).

Timbre appears to play a less central role in determining the emotional status of a piece of music. In a study of Western listeners’ perceptions of emotions expressed by Hindustani music (Balkwill and Thompson 1999), flutes were associated with peacefulness and strings with anger. In a follow-up study that compared perceptions of music from different cultures (Balkwill et al. 2004), timbre played a relatively small part in affective judgments, but a flute (vs. string) timbre was associated with sadness in Western music. There is also some evidence that soft timbres (with attenuated high frequencies) are associated with tenderness and sadness, whereas sharp timbres (with emphasized high frequencies) are associated with anger (Juslin 1997). Other musical dimensions have been examined in relation to emotion but there are too few studies to make definitive statements. Some of these dimensions include pitch height, harmonic and rhythmic complexity, specific intervals, and orchestral range (Gundlach 1935; Hevner 1935, 1936; Ilie and Thompson 2006). Previous reviews (Gabrielsson and Juslin 2003; Juslin and Laukka 2004) provide more detailed descriptions of links between musical characteristics and perceived emotion in music.

As noted, some emotions expressed by music may be interpreted correctly across cultures. Western listeners can correctly interpret joy, sadness, and anger expressed by Indian ragas (Balkwill and Thompson 1999), whereas Japanese listeners

can correctly identify these emotions expressed by Western and Hindustani music (Balkwill et al. 2004). Fritz et al. (2009) examined the perception of emotions expressed in Western music among Mafa listeners who lived in a culturally isolated region of Cameroon. A forced-choice task confirmed that these listeners could identify happiness, sadness, and fear expressed in the music at above-chance levels. These findings suggest that some affective associations with musical characteristics are present cross-culturally if not universally.

In a meta-analysis of music-performance studies, Juslin and Laukka (2003) concluded that listeners perform better than chance at interpreting happiness, sadness, anger, fear, and tenderness in music, as they do at interpreting prosodic cues in speech. They also found that anger and sadness are typically identified better than the other three emotions. One possible explanation for the discrepancy between this conclusion and the one above (re: happiness and sadness) comes from the fact that Juslin and Laukka focused on studies that used melodies as stimuli (i.e., monophonic music, such as a single voice or a trombone). Each melody was typically performed by a musician in multiple ways, with each performance designed to express a different target emotion. In other words, emotions other than happiness and sadness may be expressed and decoded successfully when the musician's specific goal is to convey a single emotion in a brief melody.

Different studies have manipulated musical characteristics in different ways. Some researchers contrasted real recordings that varied in a systematic manner (e.g., Hunter et al. 2008a), whereas others created pieces that varied along a single dimension (e.g., Gabrielsson and Juslin 1996). Such methodological differences could lead to inconsistent findings across different studies. In an attempt to facilitate comparisons across laboratories, Vieillard et al. (2008) derived a set of musical excerpts (like Eckman's faces; Ekman and Friesen 1976), with each excerpt associated primarily with one of four emotions: happiness, sadness, fear, or peacefulness. Peacefulness (not a basic emotion) was included to cover each quadrant of arousal and valence space defined by the circumplex model (Russell 1980). The happy-sounding excerpts (high arousal and positive valence) were in major mode with a fast tempo, the fearful excerpts (high arousal, negative valence) were in minor mode with some dissonance and irregular rhythms, the peaceful excerpts (low arousal, positive valence) were major with an intermediate tempo, and the sad excerpts (low arousal, negative valence) were slow and minor. Adult listeners confirmed that each excerpt was associated with its corresponding emotion. Accuracy was higher for the happy-sounding than for the other excerpts, however, and listeners sometimes misidentified peaceful-sounding excerpts as sad or happy. To derive a smaller number of more unambiguous stimuli that could be used with children, Hunter et al. (2008b) reduced the stimulus set to the five excerpts from each emotion category that were identified most reliably by adults. Compared to older listeners, younger children had more difficulty identifying sad- or peaceful-sounding excerpts correctly, but adult-like accuracy was reached by 11 years of age. The development of emotional responses to music is considered in greater detail by Trainor and Corrigan, Chap. 4.

One relatively underexplored area concerns ways in which musical characteristics *interact* in their influence on emotional responding. Such interactions are liable to be complex and at least somewhat idiosyncratic, varying from piece to piece and from listener to listener. For example, in one study, rhythm and pitch changes interacted in their influence on listeners' judgments of the emotions conveyed by melodies, and such interactions varied across stimuli even for the same emotion (Schellenberg et al. 2000).

5.3.1 Summary

The available evidence indicates that happiness and sadness are readily associated with musical characteristics such as tempo and mode, and that such associations may be evident across musical cultures. Other basic emotions such as fear and peacefulness may be perceived reliably when the stimuli are designed specifically to portray one of these emotions. Although affective associations of both tempo and mode are fairly well established, effects of other musical characteristics are poorly understood, as are the ways in which characteristics of music interact in determining the emotional status of a piece. Efforts to create standardized affective musical pieces (such as those created by Vieillard et al. 2008) may help to make results more consistent across future studies.

5.4 Measuring Emotional Responses to Music

5.4.1 Self-Reports

Most studies have used one of three methods to measure emotional responses to music. Perhaps the most common method is to ask listeners to rate the extent to which they perceive or feel a particular emotion, such as happiness (e.g., Gagnon and Peretz, 2003; Hunter et al. 2008a). Another method is to present listeners with a list of possible emotions and ask them to indicate which one (or ones) they hear (e.g., Gundlach 1935). A third approach is to require participants to rate pieces on a number of dimensions (often arousal and valence; e.g., Schmidt and Trainor 2001; Vieillard et al. 2008). Less common but more sophisticated approaches have used continuous response scales (e.g., Grewe et al. 2007a) that measure second-by-second changes in one or more dimensions of affective responding. These kinds of techniques allow specific musical events to be correlated with simultaneous changes in reported affect (Schubert 2001). In a relatively novel approach, Juslin et al. (2008) used experience sampling to demonstrate the relative frequency of various emotions in musical and non-musical contexts over long time spans. Participants were supplied with electronic devices that prompted them seven times a day to respond to a questionnaire.

All of the above methods represent different types of *self report*, which may lead to concerns of response bias and to doubts about the validity of the responses. Fortunately, people tend to be very attuned to how they are feeling (i.e., to the subjective component of their emotional responses). For example, self-reports are used to verify that mood-induction procedures are successful. The induced moods, in turn, have differential consequences for a variety of other behaviors (e.g., Husain et al. 2002; Alter and Forgas 2007; Grant et al. 2007). Self reports of participants' emotional state also correlate with *other reports* (i.e., responses provided by friends or informants; Watson and Clark 1991; Lucas et al. 1996), and they predict real-world outcomes such as suicide and longevity (Lyubomirsky et al. 2005). Indeed, the available evidence provides support for Gabrielsson's (2002, p. 128) conclusion that self reports are "the best and most natural method to study emotional responses to music."

5.4.2 *Physiological Measures*

Several researchers have attempted to measure emotional responding to music physiologically (Krumhansl 1997; Nyklíček et al. 1997; Rickard 2004; Sammler et al. 2007; Khalifa et al. 2008). Krumhansl's (1997) participants listened to music while 12 indices of physiological activity were measured. These included seven measures relating to blood flow (e.g., cardiac inter-beat-interval), three related to respiration (e.g., respiration depth), as well as skin conductance and finger temperature. The musical stimuli consisted of six excerpts that sounded happy, sad, and scary, with two excerpts representing each of the three emotions. Similarly, Nyklíček et al. (1997) measured a number of cardiac and respiratory variables while participants listened to music excerpts from one of four categories, which represented four quadrants of emotion space defined by arousal and valence (i.e., happy, agitated, sad, and serene). Both studies found differences between high- and low-arousal emotions but few differences between emotions with positive or negative valence. Although Krumhansl reported differences between responses to happy- and sad-sounding music, because happiness and sadness differ in both valence and arousal, it is difficult to determine the cause. Indeed, Nyklíček et al. reported that differences in valence accounted for only 10% of the variance in their physiological measures.

Other researchers have also found differences between responses to happy- and sad-sounding pieces on some measures (diastolic blood pressure, skin conductance, finger temperature, and zygomatic activity), but again, because these emotions differ on both valence and arousal the findings could be attributed to differences in arousal, not valence (Khalifa et al. 2008; Lundqvist et al. 2009). Moreover, reliable changes in skin-conduction responses that are evident in response to unexpected musical events are likely to be a consequence of increases in arousal (Koelsch et al. 2008b). Thus far, then, physiological measures are good at measuring differences in levels of arousal that occur in response to music listening, yet they are relatively

insensitive at discriminating responses that differ in valence. One exception involves musical stimuli that are manipulated electronically to sound pleasant (consonant) or extremely unpleasant (highly dissonant). In this instance, unpleasant stimuli lead to decreases in heart rate (Sammler et al. 2007).

One physiological measure that holds some promise is facial electromyography (EMG). EMG measurements of zygomatic (cheek) and corrugator (brow) facial muscles are associated with processing positive and negative events, respectively (Schwartz et al. 1980; Tassinary et al. 1989; Lang 1995). Witvliet and Vrana (2007) measured muscle activity while participants listened to music that was selected to fit in one of the four quadrants of the arousal and valence space. They reported greater zygomatic activity in response to pieces that elicited high arousal and positive valence, whereas corrugator activity was exaggerated in response to pieces with negative valence, regardless of arousal. Thus, corrugator activity might serve as a measure of valence. Unfortunately, *liking* ratings were lower for negatively valenced pieces, which makes it unclear whether the observed corrugator effect was a result of the negative emotion of the piece, or the fact that listeners did not like it. Indeed, although EMG is often considered to be a measure of valence, many of the relevant studies used stimuli (e.g., negative pictures; Lang 1995) that evoke both negative emotions and unpleasant evaluations. This issue is particularly relevant for music, which can evoke or express a negative emotion (e.g., sadness) yet the listener may still find the particular piece to be pleasant or beautiful.

Other researchers found few differences in corrugator responses to happy- and sad-sounding pieces, but zygomatic activity was greater for happy-sounding pieces (Khalfa et al. 2008; Lundqvist et al. 2009). In short, although the findings are equivocal about corrugator activity, increases in zygomatic activity are evident in response to happy-sounding (high arousal and positive valence) pieces. Nonetheless, fearful stimuli also increase zygomatic activity (Schwartz et al. 1980; Stemmler et al. 2001; Pauls and Stemmler 2003), which rules out a simple valence explanation of such activity.

5.4.3 Measures of Brain Activation

Several studies have used measures of brain activity to examine emotional responses to music. One approach has been to measure asymmetry in frontal electroencephalographic (EEG) activity. Greater activity in the left-frontal region is often assumed to be associated with positive affect, whereas greater activity in the right region is associated with negative affect (Davidson 1998). In line with this view, Schmidt and Trainor (2001) found greater left- and right-hemisphere activity during listening to music with positive and negative valence, respectively. Other results indicate, however, that frontal asymmetry actually measures motivational direction (i.e., approach and avoidance tendencies) rather than valence. For example, anger – a negatively valenced *approach* emotion – elicits left frontal activity (Harmon-Jones and Allen 1998; Harmon-Jones and Sigelman 2001). Others studies have reported that

(1) liked music evokes left frontal activation, whereas disliked music evokes right but slightly more bilateral activation (Altenmüller et al. 2002), (2) pleasant sounding (consonant) music evokes greater midline activity compared to unpleasant sounding (dissonant) music (Sammler et al. 2007), and (3) EEG activity varies in response to expressive and unexpressive musical performances (Koelsch et al. 2008b).

Neuroimaging approaches have had some success as measures of emotional responding (for a review, see Peretz 2010) but they too are not without problems. Several studies compared responses to pleasant and unpleasant music, but very few compared responses to music that expresses more specific emotions, such as happiness or sadness. In one exception, Green et al. (2008) used functional magnetic resonance imaging (fMRI) to compare brain activity among participants listening to novel melodies composed with major or minor scales. A pilot study confirmed that the major melodies sounded happier than the minor melodies, even though the two sets were equated for tempo. The anterior cingulate, the left parahippocampal gyrus, and the left medial frontal cortex were more active while participants listened to minor than to major melodies. The behavioral task in the scanner involved making liking judgments for each melody, however, so it is unclear whether participants actually felt happiness or sadness in response to the melodies. Moreover, although liking judgments did not differ significantly for the two sets of melodies, the observed difference between sets (i.e., minor melodies liked slightly more, $p < 0.1$) would have been significant with a slightly larger sample. Hence, differential brain-activation patterns may have stemmed from differences in liking, differences in perceived happiness and sadness, and/or differences in felt happiness and sadness.

In another study, Mitterschiffthaler et al. (2007) used fMRI recordings while listeners heard happy-, sad-, and neutral-sounding excerpts from familiar orchestral pieces (e.g., *The Blue Danube* by Johann Strauss). After each piece, participants rated whether their emotional state became happier or sadder while listening to the piece. The researchers made no attempt to measure or equate liking for the pieces, however, and the results were quite different from those of Green et al. (2008). Compared to the neutral pieces, the happy-sounding pieces elicited more activity in the parahippocampal gyrus, anterior cingulate, and ventral and dorsal striatum. For sad-sounding music, greater activity was evident in the hippocampus/amygdala region. These are many of the same regions that were implicated in studies comparing pleasant and unpleasant stimuli, which highlights the importance of equating pieces in terms of pleasantness or liking. At present, conflicting findings and the use of different behavioral tasks (e.g., liking or happy/sad ratings) and stimuli (e.g., novel melodies or familiar orchestral pieces) preclude unequivocal interpretations of the results from neuroimaging studies that measured brain-activation patterns in response to happy- and sad-sounding music.

Neuroimaging techniques are more successful, however, at distinguishing between positive (e.g., liking, pleasantness; Blood and Zatorre 2001) and negative (e.g., disliking, unpleasantness; Blood et al. 1999) responding. For example, Koelsch et al. (2006) used fMRI to record brain activity while their participants listened to pleasant (consonant) and unpleasant (dissonant) music. Pleasant and unpleasant

music activated regions related to positive and negative affect processing, respectively (e.g., the ventral striatum and the parahippocampal gyrus). These activation patterns were similar to those reported previously (Blood et al. 1999).

In addition, activity in the amygdala appears to be a marker of pleasantness. That is, reduced activity occurs in response to pleasant sounding music, whereas activity increases in response to unpleasant music (Blood and Zatorre 2001; Brown et al. 2004), even in response to a single unexpected and unpleasant sounding chord (Koelsch et al. 2008a). Brown et al. (2004) observed activation throughout the brain during music listening, but the control condition involved no listening experience – participants simply rested in the scanner. Thus, observed activation patterns may not have been specific to music or to affective responding. One method (i.e., cytoarchitectonically defined probabilistic maps) used in conjunction with fMRI allows for higher resolution images, which reveal that *both* pleasant (consonant) and unpleasant (dissonant) musical stimuli lead to enhanced activity in separate areas of the amygdala (Ball et al. 2007).

5.4.4 Patients with Brain Damage

Studies of patients with brain damage are also informative about regions that subserve the perception of specific emotions. Gosselin et al. (2005, 2007) studied the effect of temporal lobe lesions and amygdala damage on the recognition of happy, sad, peaceful, and scary emotions expressed in music. Amygdala damage is known to be associated with specific impairments in the recognition of threat signals, especially from faces (Adolphs et al. 2005). In one case study (Gosselin et al. 2007), a patient with specific bilateral amygdala lesions showed deficits in recognizing scary and sad emotions in music. Thus, the amygdala seems essential for the perception of threat from scary-sounding music, although it may also be important for sadness, which would be consistent with Mitterschiffthaler et al. (2007)'s finding of amygdala activity while listening to sad-sounding music. In another study, lesions in the temporal lobe were associated with impairments in the recognition of scary and, to a lesser extent, peaceful emotions expressed musically (Gosselin et al. 2005). It is unclear to what extent these deficits extend to *felt* emotions.

5.4.5 Summary

This section focused on four categories of measurement, each with its own strengths and weaknesses. Behavioral measures (such as scale ratings and continuous response scales) rely typically on self report yet they are reliable and valid. They also benefit from ease of interpretation, at least for subjective feelings, and they are by far the least expensive method. Physiological measures (e.g., heart rate, blood pressure) are reliable at differentiating pieces that elicit varying levels of arousal,

but not valence. Facial EMG might be an exception, although further studies using well-controlled stimuli are necessary. Neurological measures have been used successfully for differentiating responses to stimuli that differ in pleasantness, but there are no studies that examined specific emotional responses (e.g., happiness and sadness) while controlling for stimulus differences in liking or pleasantness. Lastly, research with patients suffering from brain damage has highlighted the importance of the amygdala in recognizing certain emotions (e.g., fear) in music. More studies are needed to determine whether amygdala activity extends to felt emotions, and whether other types of neurological damage impair emotional responding to music.

5.5 Perceived and Felt Emotions: Similarities and Differences

If we accept that listeners perceive emotions conveyed by music *and* respond emotionally to music, a logical question arises: how are these perceptions and feelings similar and different? In one questionnaire survey, most respondents ($\approx 70\%$) reported that when they perceive an emotion expressed musically, they often feel the same emotion (Juslin and Laukka 2004). In many studies, however, it is unclear whether listeners responded according to how the music sounds or how it makes them feel. For example, in Hevner's studies (1936, 1937), listeners were asked to choose which emotion best "fit" the music. Many other studies used similarly ambiguous response formats (e.g., Gundlach 1935; Cunningham and Sterling 1988; Kastner and Crowder 1990; Terwogt and Van Grinsven 1991). When response formats are unclear, participants may respond according to their perception of emotion expressed in the music, or to the emotion they feel in response to the music. These two constructs do not necessarily vary in tandem (see Gabrielsson 2002). For example, listeners may have no emotional response to sad-sounding music when they are in a happy mood, yet they might still recognize that a piece sounds sad. Similarly, a piece that is obviously happy-sounding may elicit negative feelings if it is associated with a negative event that happened in the past (e.g., the breakup of a relationship).

To determine the extent to which feelings differ from perceptions, it is necessary to measure both kinds of responses. A small number of studies have done this, but the results tend to be inconclusive for various reasons. In one study, listeners heard short pieces created electronically using MIDI (Vieillard et al. 2008, Experiment 1). Some listeners rated their felt emotion; others rated the emotion they perceived. Ratings were highly correlated across pieces, but feeling ratings tended to be higher than perceptions of emotion, which seems counter-intuitive. Indeed, because the pieces were unfamiliar and relatively unexpressive, one would expect actual emotional responding to be weaker than perceptions. The problem may have stemmed from asking listeners to rate either feelings *or* perceptions, when the question of interest is whether these ratings covary within listeners. Zentner et al. (2008) also examined feelings and perceptions, but they did so by asking participants to rate the general frequency with which they feel or perceive particular emotions when

listening to their favorite music. Perceptions were more common than feelings for most emotions, but this finding does not tell us whether they correlate in terms of magnitude.

Kallinen and Ravaja (2006) used 1-min musical excerpts to compare felt and perceived emotions. Listeners made 16 ratings related to their feelings as well as 16 ratings related to the emotions they perceived. Feeling and perception ratings were correlated, but perceptions tended to be stronger than feelings on ratings of arousal and activation, whereas the reverse was true for ratings of pleasantness. It is interesting to note that some pieces expressing negative emotions were rated as eliciting pleasant feelings. One problem with this study is the total number of ratings that listeners were required to make. Indeed, by the time respondents made their 32nd rating, their memory for how the piece sounded and how it made them feel may have faded. Other studies have provided converging evidence that perceptions tend to be stronger than actual feelings when both responses are measured identically (Schubert 2007a, b; Evans and Schubert 2008). Because listeners rated a small number of pieces (from 2 to 5) in each instance, however, considerations of how the magnitude of feeling and perception responses co-vary within individual listeners were precluded. In one study, feeling responses were more variable across listeners than perceiving judgments, presumably because the listeners were familiar with the stimuli (pieces from the Romantic repertoire), which may have evoked personal associations that were specific to individual listeners (Schubert 2007b).

In an attempt to address some of these methodological issues, Hunter et al. (2010) asked listeners to rate only their felt and perceived happiness and sadness in response to 32 musical stimuli. The stimuli were 30-s MIDI versions of Bach preludes that were manipulated to vary in tempo and mode. As with Kallinen and Ravaja, the perception and feeling ratings were correlated positively but not perfectly (see also Evans and Schubert 2008), and perceptions were stronger than feelings for both happy and sad ratings, as one would expect. Mediation analyses revealed that feelings were mediated by perceptions, but that the reverse was not true. In other words, when listeners responded emotionally to music, they also tended to be aware of how the music sounded. By contrast, in some instances listeners perceived emotions expressed musically without actually feeling anything.

5.5.1 *Summary*

Links between feeling and perceiving musical emotions have been understudied. Current evidence suggests that feeling and perceiving responses are correlated positively but imperfectly, and that perceptions are typically stronger than feelings. Further research in this area may clarify the mechanisms through which music evokes emotions (Juslin and Västfjäll 2008). For example, the emotional contagion hypothesis suggests that the emotion perceived in music should tend to evoke the same or a similar emotion in the listener, although not necessarily at the same intensity.

5.6 Chills

Goldstein (1980) was the first to study empirically the phenomenon of chills (or *thrills*) in response to music. He described chills as a tingling sensation resulting from a strong emotional experience. Questionnaires given to three groups of participants (university employees, medical students, and music students) indicated that about half of the population had experienced music-related chills. Chills were also reported to be fairly frequent, at least among music students, 60% of whom said they had felt them in the past week. These estimates vary, however, depending on the sample and the questions asked. Sloboda (1991) found that 90% of his participants, mostly professional and amateur musicians, reported feeling chills within the past 5 years. In another study, 86% of students enrolled in an introductory psychology class reported having the experience with some regularity (Panksepp 1995). Experiencing chills may also depend on certain personality factors. Indeed, one item on a widely used *Big Five* personality inventory (Costa and McCrae 1992) asks respondents whether they feel aesthetic chills. Across cultures and languages, this particular item is one of the most reliable predictors of the dimension of personality called *openness to experience* (McCrae 2007).

5.6.1 What Are Chills?

Goldstein (1980) reported that chills were felt most commonly in the back of the neck and upper spine, yet more than one-quarter (28%) of his respondents reported similar sensations in their legs. Chills also varied from being very brief, isolated experiences, to spreading sensations of longer duration. In one study, musical passages that elicited chills also led to increases in skin conductance response (SCR), and, in some cases, piloerection, but not to changes in skin temperature or heart rate (Craig 2005). In another study, the association between chills and increases in SCR was evident only when listeners heard emotionally powerful music; relaxing music, arousing (but not emotional) music, and watching an emotionally powerful film scene did not have the same effect (Rickard 2004).

In a positron emission tomography (PET) study, Blood and Zatorre (2001) examined brain activity while participants were experiencing music-induced chills. Using participant-selected music (with another participant's selection as a control stimulus), they found activation patterns similar to those involved in receiving rewards (i.e., greater activity in ventral striatum and dorsomedial midbrain, reduced activity in the amygdala and ventromedial prefrontal cortex). In other words, chill-inducing music appears to be something like auditory cocaine. As in other imaging studies, the authors did not consider specific emotions in their analysis. The findings might have been even stronger had the authors used different control stimuli, because there is some evidence that one listener's chill-inducing music may also

induce chills in other listeners (Panksepp 1995). Perhaps a more appropriate control would have been a second piece selected by participants that did not induce chills but was liked equally.

Self-report and brain data both imply that chills result from or occur in tandem with strong experiences of pleasure. In line with this view, some listeners report fewer chills after taking an opiate antagonist (Goldstein 1980). Nonetheless, another view holds that chills represent a measure of negative affect. For example, chills can be more common among females when they listen to sad-sounding pieces (Panksepp 1995), and among both males and females when they listen to slow-tempo music (Guhn et al. 2007). Chills also occur in response to aggressive film scenes (Geen and Rakosky 1973), erotic stimuli (Hamrick 1974), and fear imagery (Vrana 1995). In other words, chills may stem from heightened states of emotional activation rather than any specific emotion. Rickard (2004) noted that SCR – one of the most reliable physiological correlates of chills – is one of the most sensitive measures of strong affect (Andreassi 2000).

In line with this view, Huron (2006) suggested that chills are related to the fight response that sometimes arises when animals (including humans) are faced with a threatening stimulus. Both responses involve piloerection. Cold temperatures also elicit this response, which provides a means of keeping the body warm. Because piloerection makes animals appear larger, it may be an adaptive response that occurs in situations when animals need to appear to be threatening, such as when they are surprised by a potentially dangerous stimulus. If the stimulus is subsequently appraised as nonthreatening, the piloerection response may be accompanied by pleasure, similarly to the way in which other surprising events may induce pleasure.

5.6.2 *How Are Chills Elicited by Music?*

Music appears to be the most common stimulus that induces chills. For example, almost all of the participants (96%) in Goldstein's (1980) study endorsed music as a chill-inducing stimulus, although dramatic scenes (92%) and aesthetic beauty (86%) were also very common. By contrast, parades (26%) were a response option that relatively few participants selected. The association between music and chills appears to rely at least partly on familiarity. Indeed, most studies have asked participants to choose their own stimuli, namely familiar pieces that reliably induce chills (e.g., Goldstein 1980; Sloboda 1991; Blood and Zatorre 2001; Rickard 2004; Grewe et al. 2007b). Nonetheless, certain pieces seem to be relatively effective at eliciting chills across participants, such as Pink Floyd's "Post-war Dream" (Panksepp 1995) and Mozart's *Tuba Mirum* (Grewe et al. 2007b). It is impossible to rule out familiarity, however, in explaining responses to these songs. The same recording (Air Supply's 1983 hit single: "Making Love Out of Nothing at All") was found to be effective at eliciting chills in one instance but not in another. Presumably,

the song was more familiar to Americans in the 1990s (Panksepp 1995) than it was to Germans more than a decade later (Grewe et al. 2007b).

Are there musical characteristics that evoke chills with some consistency? In one report, chills occurred most commonly in response to new or unprepared harmonies and to sudden dynamic or textural changes (Sloboda 1991), which is consistent with Huron's (2006) account of chills that occur in response to unexpected but ultimately non-threatening events. In other reports, chills tended to occur when the music was slower, when the lead instrument changed, and when there was an increase in loudness (Grewe et al. 2007b; Guhn et al. 2007). The association between crescendos and chills was also reported by Panksepp (1995), although it may be evident only at certain frequencies (920–4400 Hz; Nagel et al. 2008).

5.6.3 Summary

The phenomenon of chills in response to music is complex. To date, it is clear that chills are relatively common in response to music, and that they are often a pleasurable experience associated with increases in arousal and SCR. Although chills can be evoked by unexpected changes in musical characteristics, responses seem to vary across listeners and pieces. In line with this view, one model of chills considers characteristics of the piece as well as the listener's personality and familiarity with the musical style (Grewe et al. 2007b). It remains unknown why chills arise primarily from exposure to music and other aesthetic stimuli. Experiences of pleasantness and emotional arousal may play a role, as may more complex, ambiguous feelings that music and other art forms evoke (Scherer 2004; Zentner et al. 2008), such as feelings of sadness (Panksepp 1995) that may be combined with experiences of pleasure (Hunter et al. 2008a) and neural activation of reward circuitry (Blood and Zatorre 2001).

5.7 Liking for Music

Those who argue that emotions such as happiness and sadness are not directly elicited by music (Kivy 1980; Konečni 2005; Konečni et al. 2007) tend to focus instead on the aesthetics of music. Konečni (2005, 2008) suggests that affective experience to music should be discussed in terms of a "trinity" that includes chills, being moved, and aesthetic awe, whereas Kivy (1980) believes that affective experiences in response to music consist only of enjoyment (or lack thereof). Regardless of the particular perspective, it is clear that listeners *like* some musical pieces more than others, and that any individual piece may be liked by some listeners but not by others. But what drives preferences for some pieces over others? There are bound

to be many influential factors, and a majority of these may be contextual or personal and relatively difficult to document systematically.

One recent finding points to a *contrast effect* based on the listening context: hearing a particularly good or bad musical stimulus influences evaluations of a subsequent stimulus in the opposite direction (Parker et al. 2008). Other research finds evidence that liking for music tends to be higher when listeners respond emotionally to it, and when their actual feelings parallel the emotion they perceive (Schubert 2007a; Evans and Schubert 2008). These findings are consistent with the view that people listen to music because of the way it makes them feel. Three other factors – each with multiple pieces of supporting evidence – are discussed below: consonance and dissonance, familiarity, and liking for sad-sounding music. The focus is on liking for novel pieces of music regardless of genre. Preferences for specific genres of music vary as a function of individual differences in lifestyle, social, and personality factors that are beyond the scope of the present chapter (North and Hargreaves 2007a, b, c; Rentfrow and Gosling 2003, 2006).

5.7.1 *Consonance and Dissonance*

Many music preferences may be person-specific, yet there are some preferences for features of music that appear to be widespread, and some factors that universally affect music preferences. At a very basic level is the preference for consonance over dissonance. Degree of dissonance is correlated with brain activity in regions associated with processing negative stimuli (e.g., the parahippocampal gyrus; Blood et al. 1999) and with physiological responses such as heart rate (Sammler et al. 2007). Moreover, a preference for consonance over dissonance is evident behaviorally in very young infants (Zentner and Kagan 1996, 1998; Trainor and Heinmiller 1998; Trainor et al. 2002; Trainor and Corrigan, Chap. 4). Indeed, because even 2-month olds exhibit a preference for consonance over dissonance (Trainor et al. 2002), this preference appears to be either innate or learned very early in life. In line with this view, the preference is evident cross-culturally. For example, the culturally isolated Mafa tribe from Cameroon dislike both Mafa and Western pieces when the pieces are manipulated to sound dissonant (Fritz et al. 2009). The preference for consonance may, however, be unique to humans. Other primates (i.e., tamarin monkeys) do not show a similar preference (McDermott and Hauser 2004). Future research could examine how and why a preference for consonance evolved among humans.

5.7.2 *Familiarity*

Familiarity is well established as a factor that influences preference for a stimulus. Zajonc (1968) was the first to demonstrate that simple exposure is sufficient to

manipulate degree of liking, even when participants have no explicit memory for the stimulus. This *mere exposure* effect has been replicated many times (Bornstein 1989). It has also been documented in response to music (e.g., Mull 1957; Peretz et al. 1998). Although the effect is consistent with anecdotal evidence that music becomes popular as it is played more often on the radio (Jakobovits 1966), the mere exposure effect cannot explain *disliking* for music that is overplayed, for which there is an equal amount of anecdotal evidence. Many attempts to document this inverted-U shaped function of increases followed by decreases in liking have met with only partial success (e.g., Zajonc et al. 1972).

Szpunar et al. (2004) examined effects of number of exposures on liking for music while varying both the complexity of the stimuli (from random sequences of pure tones to excerpts from orchestral recordings) and whether the exposure involved focused or incidental listening. Figure 5.2 illustrates liking ratings as a function of number of exposures, stimulus complexity, and listening condition. For pure-tone sequences, liking did not vary as a function of exposure for listeners who were required to focus on the stimuli during the exposure phase, presumably because the stimuli were aesthetically impoverished. For listeners who heard the same stimuli incidentally (i.e., in the background), however, liking increased monotonically as a function of exposure, even though listeners had no explicit memory for stimuli they had heard as many as 64 times. For this group, response patterns were a complete replication of the mere exposure effect. The same linear increase in liking as a function of

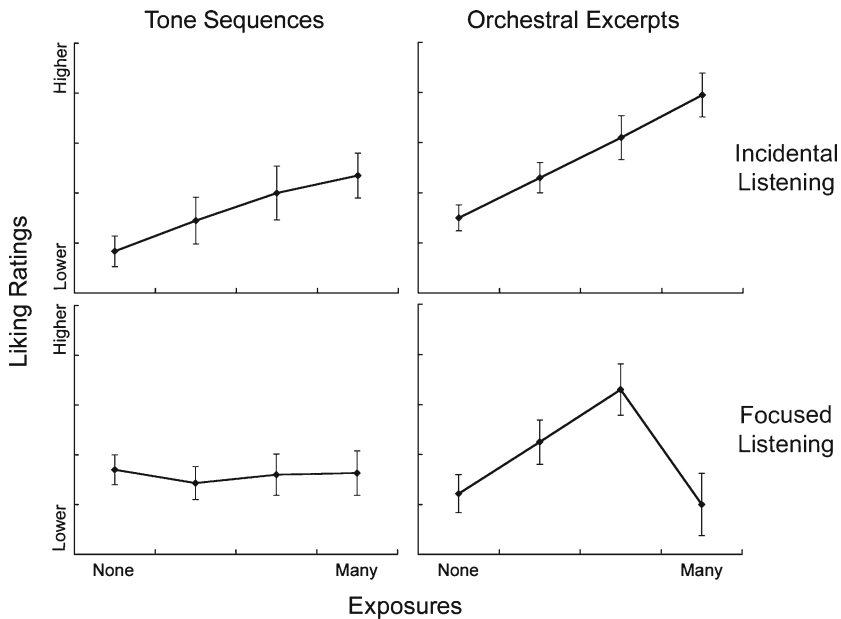


Fig. 5.2 The effect of exposure on liking as a function of number of exposures, the type of exposure, and stimulus complexity (Data are from Szpunar et al. 2004)

exposure was evident for focused listeners tested with complex stimuli (real music), except that these listeners also remembered the excerpts. Finally, the inverted-U shaped function was evident only with focused attention and more complex stimuli. For these listeners, liking and memory increased for pieces they heard twice compared to novel pieces (i.e., a baseline measure), and for pieces heard eight times compared to those heard twice. Liking for pieces heard 32 times returned to baseline levels, when memory was at ceiling.

Two main theoretical frameworks are used to account for effects of exposure on liking. The *two-factor model*, first proposed by Berlyne (1970) and developed further by Stang (1974), posits that an inverted-U shaped function is the result of the *arousal potential* of a stimulus, which should be neither too great nor too small. Unfamiliar stimuli are potential threats, which makes their arousal potential too great. With increasing familiarity and exposure to the stimulus that does not have adverse consequence, its arousal potential decreases to an optimal level and, as a result, liking for the stimulus increases. This first (increasing) part of the curve is related to Zajonc's (1968) explanation of the mere exposure effect: humans have an initial disposition to distrust an unfamiliar stimulus, but with exposure that is not harmful they begin to trust and therefore like it. The two-factor model also notes that after many exposures, boredom or fatigue sets in as the arousal potential of the stimulus decreases to less than optimal levels. This second factor accounts for decreases in liking for over-familiar songs.

Another explanation is Bornstein's (1992; Bornstein and D'Agostino 1994) *perceptual fluency/attribution model*. It suggests that previous exposure to a stimulus can result in a context-free representation of that stimulus, which increases processing fluency (i.e., speed and efficiency of processing) for subsequent exposures to the same stimulus. When the perceiver has no explicit memory for the stimulus, this fluency can be misattributed as a positive disposition towards the stimulus. An updated but similar model holds that fluency itself is pleasurable (Reber et al. 2004), which helps to explain preferences for prototypical and symmetric stimuli as well as preferences for familiar stimuli. Any factor that increases fluency should also increase liking. In both models, fluency is pleasurable when it is unexpected. Presumably, after a small number of exposures the perceiver may have the pleasant experience of unexpected fluency yet no explicit memory for the stimulus. With a greater number of exposures, the stimulus becomes familiar such that processing fluency is expected and no longer pleasurable. In short, both models suggest that when an individual is aware of the cause of fluency or expects fluency, liking should decrease. In other words, once the stimulus is remembered explicitly, participants should start to dislike it. In many instances, however, listeners both remember and like musical stimuli (Szpunar et al. 2004; Schellenberg et al. 2008). Similar findings are evident in other domains, such as when participants consciously remember and like visual stimuli (e.g., polygons or photographs of faces; Newell and Shanks 2007).

The available data on exposures, liking, and memory for music indicate that which model is appropriate depends on stimulus complexity and the listening experience. Indeed, a complete account of the data appears to require a hybrid of the two-factor and perceptual fluency/attributional models.

5.7.3 *Liking for Sad-Sounding Music*

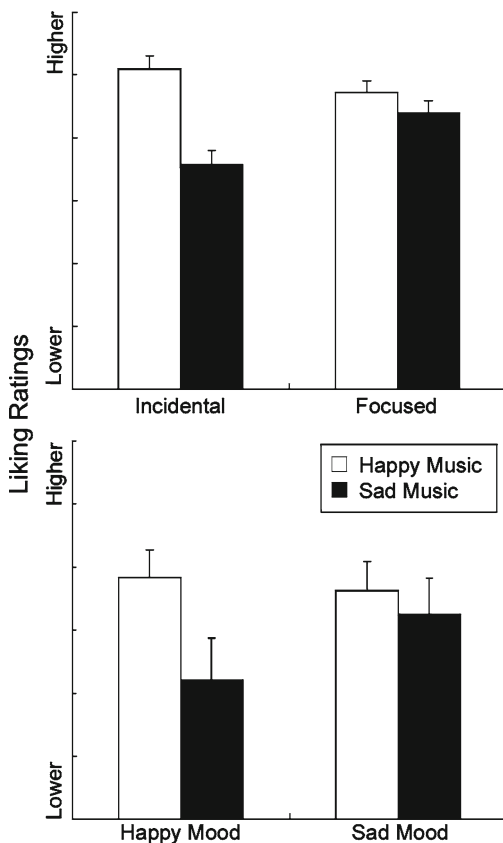
One of the most intriguing yet under-researched questions concerns sad-sounding music: why listeners like it and why it even exists. If sad-sounding music elicits sad emotions or moods, and if sadness is an unpleasant state, then why would anyone produce it or listen to it? According to most theories of emotions, including appraisal theory (Smith et al. 1993) and the circumplex model (Russell 1980), listeners should not like sad music. Indeed, participants in experiments typically prefer happy- over sad-sounding music, or they judge sad-sounding music to be unpleasant (Thompson et al. 2001; Hunter et al. 2008a; Vieillard et al. 2008). There is also evidence that listening to sad-sounding music activates a right-sided frontal asymmetry in EEG activity (Schmidt and Trainor 2001), which is a marker of avoidance motivation.

Studies of mixed feelings, which ask listeners to rate happy and sad feelings separately on two unipolar scales (Hunter et al. 2008a, 2010) indicate that listeners are actually ambivalent toward sad-sounding music. Although they report feeling sad, they also report some happy feelings as well. By contrast, happy-sounding music elicits only happy feelings. Sad-sounding music also leads to ambiguous responses when listeners are asked to make ratings of pleasantness and liking. Mixed feelings cannot fully explain listening to sad-sounding music, however, because when given the choice between a stimulus that elicits just happiness and one that elicits both happiness and sadness, an individual should prefer the former.

Schellenberg et al. (2008) examined the effect of repeated exposures on liking for happy- and sad-sounding music. The conditions were similar to those used by Szpunar et al. (2004). Listeners rated their liking for novel pieces as well as for pieces heard 2, 8, and 32 times previously. The initial exposures occurred during focused or incidental listening. For focused listeners, the typical preference for happy music was evident (i.e., liking ratings were significantly higher for happy music). For incidental listeners, however, happy and sad pieces were liked equally (see Fig. 5.3, upper panel). Because the incidental listeners completed a demanding and lengthy distractor task during the exposure phase, the sad-sounding music they heard later might have had a pleasant calming effect. Alternatively (or in conjunction), the demanding and lengthy task may have put listeners in a negative mood. In other words, the appeal of sad-sounding music might increase when listeners are in a negative mood.

The latter hypothesis was tested directly in a follow-up study (Hunter and Schellenberg 2008). The authors compared liking ratings for happy- and sad-sounding music after happy, sad, and neutral mood inductions. To induce a mood, participants viewed a set of pictures. They then wrote a short paragraph about their emotional response to one of the pictures. Responses included emotional memories, issues participants felt strongly about, or just a description of the picture (if no emotion was evoked). As one would expect, the typical preference for happy- over sad-sounding music was evident after happy and neutral mood inductions. When a sad mood was induced, however, happy- and sad-sounding pieces were liked

Fig. 5.3 Liking for happy-and-sad-sounding music as a function of the type of previous exposure (*upper panel*, data from Schellenberg et al. 2008) and listener's mood (*lower panel*, data from Hunter and Schellenberg 2008)



equally (Fig. 5.3, lower panel). Because levels of liking for happy-sounding music were virtually identical after happy and sad mood inductions, it is safe to conclude that sad moods increased the appeal of sad-sounding music. Sad-sounding music may also have a cathartic effect by helping listeners to get in touch with their sad feelings, thereby allowing these feelings to dissipate.

5.7.4 Summary

This section discussed a number of influences on liking for music. Consonance appears to have a fairly basic influence on liking. It is typically preferred to dissonance regardless of age and familiarity with Western musical traditions. Moreover, such a preference may be uniquely human. Familiarity is another basic influence on liking. Although initial exposures to music typically increase liking, over-familiarity often leads to disliking. The association between exposures and liking varies, however,

as a function of stimulus complexity and the listening experience. Finally, listeners tend to prefer happy- over sad-sounding music, yet they often choose to listen to sad-sounding music. The appeal of sad-sounding music appears to increase when listeners are fatigued or sad.

5.8 Conclusions and Future Directions

Several themes emerge from this review of research on links between music and emotion. One is the still unanswered question of the nature of affective reactions to music: whether they consist of true emotions, moods, aesthetic emotions, or liking responses. Musical emotions appear to differ from everyday emotions, at least as defined by appraisal theory. Emotional reactions to music (with a few exceptions, e.g., anger because a particularly disliked song is playing) are not the same as those to everyday events because the stimulus (music) is not goal-relevant and does not elicit the usual cognitive appraisals. Nevertheless, there may be other routes for activating emotions through music (Juslin and Västfjäll 2008). Another possibility is that affective responses to music are short-lived moods rather than emotions, which are similar to those evoked by nonmusical stimuli. Affective responses to music could also represent a separate class of emotions that consists primarily of the subjective feeling and may be specific to aesthetic stimuli (Scherer 2004; Zentner et al. 2008).

Thus, if we accept that listeners respond affectively to music, we have three possible explanations: (1) listeners respond emotionally but without cognitive appraisals, (2) music evokes temporary changes in mood, and (3) music evokes a special class of aesthetic emotions. These three explanations for affective reactions to music are not mutually exclusive. Aesthetic emotions may be one particular type of mood or emotion and it is not obvious how these alternatives could be distinguished empirically. The extent to which moods, emotions, or aesthetic emotions are elicited could, however, depend on the listener, the piece, and the context. Future research on individual differences and the effect of the context on the nature of affective reactions to music may lead to insights that could help to resolve theoretical debates, which are currently deadlocked.

Another issue is the ubiquity of the use of the circumplex model in research on music and emotion and the need to consider alternatives in some instances. Recent evidence suggests that this two-dimensional model is not appropriate for every context. More specifically, emotional responses may vary along dimensions other than valence and arousal (Fontaine et al. 2007). Further studies could focus on validating these additional two dimensions with musical stimuli. Moreover, findings of mixed feelings (i.e., coactivated positive and negative affect) in response to music (Hunter et al. 2008a, 2010) and nonmusical stimuli (e.g., Larsen et al. 2004) suggest that valence is better conceptualized as two dimensions rather than one. If a particular research question is concerned only with certain emotions, however, or if the method is unlikely to evoke feelings of ambiguity, then the circumplex

model may provide a parsimonious explanation of response patterns. When the question involves different emotions that lie in the same quadrant of the circumplex, or when emotional coactivation and mixed feelings are likely to arise (e.g., when examining responses to sad-sounding music), alternatives to this model should be considered. Measuring positive and negative valence separately simply requires the use of two rating scales rather than one, or, alternatively, a two-dimensional response grid (Larsen et al. 2009). If reliable neural or physiological correlates of positive and negative affect were to be identified at some point in the future, one prediction is that they would be coactivated when listeners experience mixed feelings in response to music.

Although the affective correlates of tempo and mode are well established, research to date on other musical dimensions is inconsistent and suffers from a diversity of methods and stimuli. A second problem is that these dimensions are often considered in isolation of one another. One avenue for future studies would be to consider in more detail how these dimensions interact (Schellenberg et al. 2000). For example, is the effect of loudness on emotional responding independent of whether the piece has a fast or slow tempo? If not, for which emotions are interactions evident? Further use of continuous response measures (Schubert 2001; Grewe et al. 2007a) could also advance our understanding of the links between musical characteristics and emotional responding, which may change from moment to moment while listening to a single piece.

It is also important to separate effects of liking and pleasantness from more specific emotional responses such as happiness and sadness. This problem is particularly chronic for studies that use physiological and neurological measures of responding to music. Obviously, the most solid findings arise when multiple methods of measurement lead to the same conclusions. To date, behavioral methods (self reports) have been remarkably effective at finding consistency for some affective responses to music. For example, fast music in a major key is rated as happy, in terms of the listener's perceptions and their feelings and by different populations across time and contexts. The reliability of self reports as a measure of emotional responding is supported further by evidence from mood studies, comparisons of self and other reports, and real-world predictive validity (e.g., Lucas et al. 1996; Husain et al. 2002; Lyubomirsky et al. 2005).

Unfortunately, psychologists' infatuation with physiological and neuroimaging techniques has come at a cost, both literally, in terms of the funding required, and figuratively, in terms of equivocal findings that differ from one study to the next. Thus far, physiological methods have difficulty distinguishing responses that differ in valence, whereas neuroimaging techniques identify emotional positivity (i.e., approaching, liking, pleasantness) versus negativity (i.e., avoiding, disliking, unpleasantness) with a modicum of consistency, yet they remain insensitive to distinctions between more specific emotions, such as happiness and sadness. Clear conclusions about differences in neurological responding to happy-sounding (e.g., major key) and sad-sounding (e.g., minor key) musical pieces require equating the pieces for liking and pleasantness. Evidence of consistent differences in physiological and neurological activity when listening to happy- and sad-sounding

(but equally pleasing) music would provide additional support for claims that affective reactions to music are not merely aesthetic evaluations (Konečni 2008) or liking responses (Kivy 1980).

Careful distinction between measuring felt and perceived emotions is also crucial to a complete understanding of associations between music and emotion. Researchers need to be clear about which variable is the one of interest. As some studies have shown (Kallinen and Ravaja 2006; Schubert 2007a, b; Evans and Schubert 2008; Zentner et al. 2008; Hunter et al. 2010), feelings do not necessarily vary in tandem with perceptions. Further studies could seek to elucidate conditions under which they diverge. Again, some of these conditions are likely to be idiosyncratic (e.g., an individual's emotional associations with a particular song), whereas others might be documented relatively easily in a laboratory setting (e.g., by manipulating listeners' moods).

A separate issue about which much is still unknown concerns chills, which are the strongest feelings induced by music. Researchers have been able to describe the phenomenological nature of chills and to measure physiological and brain responses, but attempts to link chills to musical events have been less successful. Chills appear to be elicited by changes in musical structure, but the nature of the change varies across studies. One possible next step would be to test models of chills that take into account the person, the stimulus, and the context (Grewe et al. 2007b). Further research could also seek to examine chills in response to stimuli other than music, in order to determine similarities and differences in chill responses across domains. A better understanding of the adaptive relevance of the chill response would also be most welcome.

Liking is perhaps better understood and more predictable than other emotional responses to music. Outstanding issues include a more complete understanding of individual differences in liking for music as a function of exposure, and how, for example, preexisting music preferences for some genres might accelerate or delay the increases and decreases in liking that have been documented. Moreover, much remains to be discovered about liking for sad-sounding music. Preliminary findings suggest that listeners warm up to sad-sounding music when they are fatigued or in sad mood (Hunter and Schellenberg 2008; Schellenberg et al. 2008). Future research could explore further the links between the listeners' emotional state and liking for sad-sounding music, and the possible role of sad-sounding music in catharsis and mood repair. Finally, although manipulations have been identified that *equate* liking for happy- and sad-sound music, it would be informative to document instances when listeners actually *prefer* sad- over happy-sounding music.

Other issues not dealt with in this chapter also bear consideration. Many scholars have suggested evolutionary explanations for the role of music in human life that are affect-related, including emotional communication (Juslin 2001), synchronicity and social cohesion (Brown 2000), and attracting a mate (Darwin 1871; Miller 2000). One possibility is that aesthetic emotional responses are linked to natural and sexual selection. It is also important to consider listeners and musical genres from different cultures. The most useful models of affective responses to music should be applicable across different musical traditions and different groups of listeners.

Acknowledgments This work was supported by the Social Sciences and Engineering Research Council of Canada.

References

- Adolphs R, Gosselin F, Buchanan TW, Tranel D, Schyns P, Damasio AR (2005) A mechanism for impaired fear recognition after amygdala damage. *Nature* 433:68–72.
- Altenmüller E, Schürmann K, Lim VK, Parlitz D (2002) Hits to the left flops to the right: different emotions during listening to music are reflected in cortical lateralization patterns. *Neuropsychologia* 40:2242–2256.
- Alter AL, Forgas JP (2007) On being happy but fearing failure: the effects of mood on self-handicapping strategies. *J Exp Soc Psychol* 43:947–954.
- Andreassi JL (2000) *Psychophysiology: Human Behaviour and Physiological Response*, 4th ed. Mahwah, NJ: Lawrence Erlbaum.
- Balkwill L-L, Thompson WF (1999) A cross-cultural investigation of the perception of emotion in music: psychophysical and cultural cues. *Music Percept* 17:43–64.
- Balkwill L-L, Thompson WF, Matsunaga R (2004) Recognition of emotion in Japanese Western and Hindustani music by Japanese listeners. *Jpn Psychol Res* 46:337–349.
- Ball T, Rahm B, Eickhoff SB, Schulze-Bonhage A, Speck O, Mutschler I (2007) Response properties of human amygdala subregions: evidence based on functional MRI combined with probabilistic anatomical maps. *PLoS One* 2(3):e307.
- Berlyne DE (1970) Novelty, complexity and hedonic value. *Percept Psychophys* 8:279–286.
- Bigand E, Vieillard S, Madurell F, Marozeau J, Dacquet A (2005) Multidimensional scaling of emotional response to music: the effect of musical expertise and of the duration of the excerpts. *Cogn Emotion* 19:1113–1139.
- Blood AJ, Zatorre RJ (2001) Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proc Natl Acad Sci USA* 98:11818–11823.
- Blood AJ, Zatorre RJ, Bermudez P, Evans AC (1999) Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions. *Nat Neurosci* 2:382–387.
- Bornstein RF (1989) Exposure and affect: overview and meta-analysis of research 1968–1987. *Psychol Bull* 106:265–289.
- Bornstein RF (1992) Inhibitory effects of awareness on affective responding. In Clark MS (ed), *Emotion: Review of Personality and Social Psychology* (No. 13). Thousand Oaks, CA: Sage, pp. 235–255.
- Bornstein RF, D’Agostino PR (1994) The attribution and discounting of perceptual fluency: preliminary tests of a perceptual fluency/attributional model of the mere exposure effect. *Soc Cogn* 12:103–128.
- Brown S (2000) The “musilanguage” model of music evolution. In Wallin N, Merker B, Brown S (eds), *The Origins of Music*. Cambridge, MA: MIT Press, pp. 271–300.
- Brown S, Martinez MJ, Parsons LM (2004) Passive music listening spontaneously engages limbic and paralimbic systems. *NeuroReport* 15:2033–2037.
- Cacioppo JT, Berntson GG (1994) Relationship between attitudes and evaluative space: a critical review with emphasis on the separability of positive and negative substrates. *Psychol Bull* 115:401–423.
- Clore GL, Schwarz N, Conway M (1994) Affective causes and consequences of social information processing. In Wyer RS, Srull TK (eds), *Handbook of Social Cognition*. Hillsdale, NJ: Erlbaum, pp. 323–417.
- Costa PT Jr, McCrae RR (1992) *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) Professional Manual*. Odessa, FL: Psychological Assessment Resources.

- Craig DG (2005) An exploratory study of physiological changes during “chills” induced by music. *Musicae Scientiae* 9:273–287.
- Cunningham JG, Sterling RS (1988) Developmental change in the understanding of affective meaning in music. *Motiv Emotion* 12:399–413.
- Damasio AR, Grabowski TJ, Bechara A, Damasio H, Ponto LLB, Parvizi J, Hichwa RD (2000) Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nat Neurosci* 3:1049–1056.
- Darwin, C. (1871) *The Descent of Man, and Selection in Relation to Sex* (2 vols.). London: Murray.
- Davidson RJ (1998) Anterior electrophysiological asymmetries emotion and depression: conceptual and methodological conundrums. *Psychophysiology* 35:607–614.
- Davies S (2001) Philosophical perspectives on music’s expressiveness. In Juslin PN, Sloboda JA (eds), *Music and Emotion: Theory and Research*. Oxford: Oxford University Press, pp. 23–44.
- Diener E, Iran-Nejad A (1986) The relationship in experience between various types of affect. *J Pers Soc Psychol* 50:1031–1038.
- Ekman P (1984) Expression and the nature of emotion. In Scherer KR, Ekman P (eds), *Approaches to Emotion*. Hillsdale, NJ: Erlbaum, pp. 319–344.
- Ekman P (1992) Are there basic emotions? *Psychol Rev* 99:550–553.
- Ekman P, Friesen WV (1976) *Pictures of Facial Affect*. Palo Alto, CA: Consulting Psychologists Press.
- Evans P, Schubert E (2008) Relationships between expressed and felt emotions in music. *Musicae Scientiae* 12:75–99.
- Fontaine JRJ, Scherer KR, Roesch EB, Ellsworth PC (2007) The world of emotions is not two-dimensional. *Psychol Sci* 18:1050–1057.
- Frijda NH (1993) Moods, emotion episodes, and emotions. In Lewis M, Haviland JM (eds), *Handbook of Emotions*. New York: Guilford Press, pp. 381–403.
- Frijda NH (1994) Varieties of affect: emotions and episodes moods and sentiments. In Ekman P, Davidson RJ (eds), *The Nature of Emotion: Fundamental Questions*. New York: Oxford University Press, pp. 59–67.
- Fritz T, Jentscke S, Gosselin N, Sammler D, Peretz I, Turner R, Friederici AD, Koelsch S (2009) Universal recognition of three basic emotions in music. *Curr Biol* 19: 573–576.
- Gabrielsson A (2002) Emotion perceived and emotion felt: same or different? *Musicae Scientiae* (Special issue 2001–2002):123–147
- Gabrielsson A, Juslin PN (1996) Emotional expression in music performance: between the performer’s intention and the listener’s experience. *Psychol Music* 24:68–91.
- Gabrielsson A, Juslin PN (2003) Emotional expression in music. In Davidson RJ, Scherer KR, Goldsmith HH (eds), *Handbook of Affective Sciences*. Oxford: Oxford University Press, pp. 503–534.
- Gagnon L, Peretz I (2003) Mode and tempo relative contributions to “happy-sad” judgments in equitone melodies. *Cogn Emotion* 17:25–40.
- Geen RG, Rakosky JJ (1973) Interpretations of observed aggression and their effect on GSR. *J Exp Res Pers* 6:280–292.
- Goldstein A (1980) Thrills in response to music and other stimuli. *Physiol Psychol* 8:126–129.
- Gosselin N, Peretz I, Noulhiane M, Hasboun D, Beckett C, Baulac M, Samson S (2005) Impaired recognition of scary music following unilateral temporal lobe excision. *Brain* 128:628–640.
- Gosselin N, Peretz I, Johnsen E, Adolphs R (2007) Amygdala damage impairs emotion recognition from music. *Neuropsychologia* 45:236–244.
- Grant VV, Stewart SH, Birch CD (2007) Impact of positive and anxious mood on implicit alcohol-related cognitions in internally motivated undergraduate drinkers. *Addict Behav* 32:2226–2237.
- Green AC, Børrentsen KB, Stødkilde-Jørgensen H, Wallentin M, Roepstorff A, Vuust P (2008) Music in minor activates limbic structures: a relationship with dissonance? *NeuroReport* 19:711–715.
- Grewe O, Nagel F, Reinhard K, Altenmüller E (2007a) Emotions over time: synchronicity and development of subjective, physiological, and facial affective responses to music. *Emotion* 7:774–788.

- Grewe O, Nagel F, Reinhard K, Altenmüller E (2007b) Listening to music as a re-creative process: physiological psychological and psychoacoustical correlates of chills and strong emotions. *Music Percept* 24:297–314.
- Guhn M, Hamm A, Zentner M (2007) Physiological and musico-acoustic correlates of the chill response. *Music Percept* 24:473–483.
- Gundlach RH (1935) Factors determining the characterization of musical phrases. *Am J Psychol* 47:624–643.
- Hamrick ND (1974) Physiological and verbal responses to erotic visual stimuli in a female population. *Behav Eng* 2:2–16.
- Harmon-Jones E, Allen JB (1998) Anger and frontal brain activity: EEG asymmetry consistent with approach motivation despite negative affective valence. *J Pers Soc Psychol* 74:1310–1316.
- Harmon-Jones E, Sigelman J (2001) State anger and prefrontal brain activity: evidence that insult-related relative left-prefrontal activation is associated with experienced anger and aggression. *J Pers Soc Psychol* 80:797–803.
- Hevner K (1935) The affective character of the major and minor modes in music. *Am J Psychol* 47:103–118.
- Hevner K (1936) Experimental studies of the elements of expression in music. *Am J Psychol* 48:246–268.
- Hevner K (1937) The affective value of pitch and tempo in music. *Am J Psychol* 49:621–630.
- Hunter PG, Schellenberg EG (2008) Misery loves company: liking for sad-sounding music increases when listeners are in a sad affective state. Paper presented at the meeting of the Psychonomic Society, Chicago, November 13–16, 2008.
- Hunter PG, Schellenberg EG, Schimmack U (2008a) Mixed affective responses to music with conflicting cues. *Cogn Emotion* 22:327–352.
- Hunter PG, Schellenberg EG, Stalinski SM (2008b) Developmental changes in liking for and recognition of emotion in music. Paper presented at the Auditory Perception and Cognition Action Meeting, Chicago, November 13, 2008.
- Hunter PG, Schellenberg EG, Schimmack U (2010) Feelings and perceptions of happiness and sadness induced by music: similarities, differences, and mixed emotions. *Psychol Aesthet Creativ Arts* 4:47–56.
- Huron D (2006) *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press.
- Husain G, Thompson WF, Schellenberg EG (2002) Effects of musical tempo and mode on arousal mood and spatial abilities. *Music Percept* 20:151–171.
- Ilie G, Thompson WF (2006) A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Percept* 23:319–329.
- Ivanov VK, Geake JG (2003) The Mozart effect and primary school children. *Psychol Music* 31:405–413.
- Jakobovits IA (1966) Studies of fads: I. The “Hit Parade.” *Psychol Rep* 18:443–450.
- Juslin PN (1997) Perceived emotional expression in synthesized performances of a short melody: capturing the listener’s judgment policy. *Musicae Scientiae* 1:225–256.
- Juslin PN (2001) Communicating emotion in music performance: a review and theoretical framework. In Juslin PN, Sloboda JA (eds), *Music and Emotion: Theory and Research*. Oxford: Oxford University Press, pp. 309–337.
- Juslin PN, Laukka P (2003) Communication of emotions in vocal expression and music performance: different channels same code? *Psychol Bull* 129:770–814.
- Juslin PN, Laukka P (2004) Expression, perception, and induction of musical emotions: a review and a questionnaire study of everyday listening. *J New Music Res* 33:217–238.
- Juslin PN, Västfjäll D (2008) Emotional responses to music: the need to consider underlying mechanisms. *Behav Brain Sci* 31: 559–621.
- Juslin PN, Liljeström S, Västfjäll D, Barradas G, Silva A (2008) An experience sampling study of emotional reactions to music: listener, music, and situation. *Emotion* 5:668–683.
- Kallinen K, Ravaja N (2006) Emotion perceived and emotion felt: same and different. *Musicae Scientiae* 10:191–213.

- Kastner MP, Crowder RG (1990) Perception of the major/minor distinction: IV Emotional connotations in young children. *Music Percept* 8:189–202.
- Kellaris JJ, Rice RC (1993) The influence of tempo loudness and gender of listener on responses to music. *Psychol Market* 10:15–29.
- Khalifa S, Roy M, Rainville P, Dalla Bella S, Peretz I (2008) Role of tempo entrainment in psychophysiological differentiation of happy and sad music? *Int J Psychophysiol* 68:17–26.
- Kivy P (1980) *The Corded Shell*. Princeton, NJ: Princeton University Press.
- Kivy P (1990) *Music Alone: Philosophical Reflections on the Purely Musical Experience*. Ithaca, NY: Cornell University Press.
- Kivy P (2001) *New Essays on Musical Understanding*. Oxford: Clarendon Press.
- Koelsch S, Fritz T, Cramon DYV, Müller K, Friederici AD (2006) Investigating emotion with music: an fMRI study. *Hum Brain Mapp* 27:239–250.
- Koelsch S, Fritz T, Schlaug G (2008a) Amygdala activity can be modulated by unexpected chord functions during music listening. *NeuroReport* 19:1815–1819.
- Koelsch S, Kilches S, Steinbeis N, Schelinski S (2008b) Effects of unexpected chords and of performer's expressions on brain responses and electrodermal activity. *PLoS ONE* 3(7):e2631.
- Konečni VJ (2005) The aesthetic trinity: awe, being moved, thrills. *Bull Psychol Arts* 5:27–44.
- Konečni VJ (2008) Does music induce emotion? A theoretical and methodological analysis. *Psychol Aesthet Creativ Arts* 2:115–129.
- Konečni VJ, Wanic RA, Brown A (2007) Emotional and aesthetic antecedents and consequences of music-induced thrills. *Am J Psychol* 120:619–643.
- Kreutz G, Ott U, Teichmann D, Osawa P, Vaitl D (2008) Using music to induce emotions: influences of musical preference and absorption. *Psychol Music* 36:101–126.
- Krumhansl CL (1997) An exploratory study of musical emotions and psychophysiology. *Can J Exp Psychol* 51:336–353.
- Lang PJ (1995) The emotion probe: studies of motivation and attention. *Am Psychol* 50:372–385.
- Lang PJ, Bradley MM, Cuthbert BN (1997) *International Affective Picture System (IAPS): Technical Manual and Affective Ratings*. NIMH Center for the Study of Emotion and Attention. Gainesville, FL: University of Florida.
- Larsen JT, McGraw AP, Cacioppo JT (2001) Can people feel happy and sad at the same time? *J Pers Soc Psychol* 81:684–696.
- Larsen JT, McGraw AP, Mellers BA, Cacioppo JT (2004) The agony of victory and thrill of defeat: mixed emotional reactions to disappointing wins and relieving losses. *Psychol Sci* 15:325–330.
- Larsen JT, Norris CJ, McGraw AP, Hawley LC, Cacioppo JT (2009) The evaluative space grid: a single-item measure of positivity and negativity. *Cogn Emotion* 23:453–480.
- Levinson J (1996) *The Pleasures of Aesthetics: Philosophical Essays*. Ithaca, NY: Cornell University Press.
- Lucas RE, Diener E, Suh E (1996) Discriminant validity of well-being measures. *J Pers Soc Psychol* 71:616–628.
- Lundqvist L-O, Carlsson F, Hilmersson P, Juslin PN (2009) Emotional responses to music: experience, expression, physiology. *Psychol Music* 37:61–90.
- Lyubomirsky S, King L, Diener E (2005) The benefits of frequent positive affect: does happiness lead to success? *Psychol Bull* 131:803–855.
- Maddel G (2002) *Philosophy Music and Emotion*. Edinburgh: Edinburgh University Press.
- Matravers D (1998) *Art and Emotion*. Oxford: Clarendon Press.
- McCrae RR (2007) Aesthetic chills as a universal marker of openness to experience. *Motiv Emotion* 31:5–11.
- McDermott J, Hauser M (2004) Are consonant intervals music to their ears? Spontaneous acoustic preferences in a nonhuman primate. *Cognition* 94:B11–B21.
- Meyer LB (1956) *Emotion and Meaning in Music*. Chicago, IL: Chicago University Press.
- Miller G (2000) Evolution of human music through sexual selection. In Wallin NL, Merker B, Brown S (eds), *The Origins of Music*. Cambridge, MA: MIT Press, pp. 329–360.

- Mitterschiffthaler MT, Fu CHY, Dalton JA, Andrew CM, Williams SCR (2007) A functional MRI study of happy and sad affective states induced by classical music. *Hum Brain Mapp* 28:1150–1162.
- Morris WN (1992) A functional analysis of the role of mood in affective systems. In Clark MS (ed), *Emotion: Review of Personality and Social Psychology*. Newbury Park, CA: Sage, pp. 213–234.
- Mull HK (1957) The effect of repetition upon the enjoyment of modern music. *J Psychol* 43:155–162.
- Nagel F, Kopiez R, Grewe O, Altenmüller E (2008) Psychoacoustical correlates of musically induced chills. *Musicae Scientiae* 12:101–113.
- Nantais KM, Schellenberg EG (1999) The Mozart effect: an artifact of preference. *Psychol Sci* 10:370–373.
- Newell BR, Shanks DR (2007) Recognising what you like: examining the relation between the mere-exposure effect and recognition. *Eur J Cogn Psychol* 19:103–118.
- Niedenthal PM, Krauth-Gruber S, Ric F (2006) *Psychology of Emotion: Interpersonal Experiential and Cognitive Approaches*. New York: Psychology Press.
- North AC, Hargreaves DJ (2007a) Lifestyle correlates of musical preference: 2. Media, leisure time and music. *Psychol Music* 35:179–200.
- North AC, Hargreaves DJ (2007b) Lifestyle correlates of musical preference: 1. Relationships, living arrangements, beliefs, and crime. *Psychol Music* 35:58–87.
- North AC, Hargreaves DJ (2007c) Lifestyle correlates of musical preference: 3. Travel, money, education, employment and health. *Psychol Music* 35:473–497.
- Nyklíček I, Thayer JF, Van Doornen LJP (1997) Cardiorespiratory differentiation of musically-induced emotions. *J Psychophysiol* 11:304–321.
- Ortony A, Turner TJ (1990) What's basic about basic emotions? *Psychol Rev* 97:315–331.
- Panksepp J (1995) The emotional sources of “chills” induced by music. *Music Percept* 13:171–207.
- Parker S, Bascom J, Rabinovitz B, Zellner D (2008) Positive and negative hedonic contrast with musical stimuli. *Psychol Aesthet Creativ Arts* 2:171–174.
- Pauls CA, Stemmler G (2003) Repressive and defensive coping during fear and anger. *Emotion* 3:284–302.
- Peretz I (2010) Towards a neurobiology of musical emotions. In Juslin PN, Sloboda JA (eds), *Handbook of Music and Emotion: Theory, Research, Applications*. Oxford: Oxford University Press, pp. 99–126.
- Peretz I, Gaudreau D, Bonnel AM (1998) Exposure effects on music preferences and recognition. *Mem Cogn* 15:379–388.
- Rauscher FH, Shaw GL, Ky KN (1993) Music and spatial task performance. *Nature* 365:611.
- Reber R, Schwarz N, Winkielman P (2004) Processing fluency and aesthetic pleasure: is beauty in the perceiver's processing experience? *Pers Soc Psychol Rev* 8:364–382.
- Rentfrow PJ, Gosling SD (2003) The do re mi's of everyday life: the structure and personality correlates of music preferences. *J Pers Soc Psychol* 84: 1236–1256.
- Rentfrow PJ, Gosling SD (2006) Message in a ballad: the role of music preferences in interpersonal perception. *Psychol Sci* 17:236–242.
- Rickard NS (2004) Intense emotional response to music: a test of the physiological arousal hypothesis. *Psychol Music* 32:371–388.
- Ridley A (1995) *Music Value and the Passions*. Ithaca, NY: Cornell University Press.
- Russell JA (1980) A circumplex model of affect. *J Pers Soc Psychol* 39:1161–1178.
- Russell JA (2003) Core affect and the psychological construction of emotion. *Psychol Rev* 110:145–172.
- Russell JA, Carroll JM (1999) On the bipolarity of positive and negative affect. *Psychol Bull* 125:3–30.
- Sammner D, Grigutsch M, Fritz T, Koelsch S (2007) Music and emotion: electrophysiological correlates of the processing of pleasant and unpleasant music. *Psychophysiology* 44: 293–304.

- Schellenberg EG, Hallam S (2005) Music listening and cognitive abilities in 10- and 11-year-olds: the Blur effect. *Ann NY Acad Sci* 1060:202–209.
- Schellenberg EG, Krysciak A, Campbell RJ (2000) Perceiving emotion in melody: interactive effects of pitch and rhythm. *Music Percept* 18:155–171.
- Schellenberg EG, Nakata T, Hunter PG, Tamoto S (2007) Exposure to music and cognitive performance: tests of children and adults. *Psychol Music* 35:5–19.
- Schellenberg EG, Peretz I, Vieillard S (2008) Liking for happy-and-sad-sounding music: effects of exposure. *Cogn Emotion* 22:218–237.
- Scherer KR (2004) Which emotions can be induced by music? What are the underlying mechanisms? And how can we measure them? *J New Music Res* 33:239–251.
- Schimmack U (2001) Pleasure displeasure and mixed feelings: are semantic opposites mutually exclusive? *Cogn Emotion* 15:81–97.
- Schimmack U, Crites SL (2005) The origin and structure of affect. In Albarracín D, Johnson BT, Zanna MP (eds), *The Handbook of Attitudes*. Mahwah, NJ: Lawrence Erlbaum, pp. 397–436.
- Schmidt LA, Trainor LJ (2001) Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions. *Cogn Emotion* 15:487–500.
- Schubert E (2001) Continuous measurement of self-report emotional response to music. In Juslin PN, Sloboda JA (eds), *Music and Emotion: Theory and Research*. Oxford: Oxford University Press, pp. 393–414.
- Schubert E (2004) Modeling perceived emotion with continuous musical features. *Music Percept* 21:561–585.
- Schubert E (2007a) The influence of emotion, locus of emotion and familiarity upon preference in music. *Psychol Music* 35:499–515.
- Schubert E (2007b) Locus of emotion: the effect of task order and age on emotion perceived and emotion felt in response to music. *J Music Therapy* 44:344–368.
- Schwartz GE, Brown S-L, Ahern GL (1980) Facial muscle patterning and subjective experience during affective imagery: sex differences. *Psychophysiology* 17:75–82.
- Siemer M (2001) Mood-specific effects on appraisal and emotion judgments. *Cogn Emotion* 15:453–485.
- Siemer M (2005) Mood-congruent cognitions constitute mood experience. *Emotion* 5:296–308.
- Sloboda JA (1991) Music structure and emotional response: some empirical findings. *Psychol Music* 19:110–120.
- Smith CA, Haynes KN, Lazarus RS, Pope LK (1993) In search of the “hot” cognitions: attributions appraisals and their relation to emotion. *J Pers Soc Psychol* 65:916–929.
- Stang DJ (1974) Methodological factors in mere exposure research. *Psychol Bull* 81:1014–1025.
- Stemmler G, Heldmann M, Pauls CA, Scherer T (2001) Constraints for emotion specificity in fear and anger: the context counts. *Psychophysiology* 38:275–291.
- Szpunar KK, Schellenberg EG, Pliner P (2004) Liking and memory for musical stimuli as a function of exposure. *J Exp Psychol Learn Mem Cogn* 30:370–381.
- Tassinary LG, Cacioppo JT, Geen TR (1989) A psychometric study of surface electrode placements for facial electromyographic recording: I. The brow and cheek muscle regions. *Psychophysiology* 26:1–16.
- Terwogt MM, Van Grinsven F (1991) Musical expression of moodstates. *Psychol Music* 19:99–109.
- Thompson WF, Schellenberg EG, Husain G (2001) Arousal mood and the Mozart effect. *Psychol Sci* 12:248–251.
- Trainor LJ, Heinmiller BJ (1998) The development of evaluative responses to music: infants prefer to listen to consonance over dissonance. *Infant Behav Dev* 21:77–88.
- Trainor LJ, Tsang CD, Cheung VHW (2002) Preference for sensory consonance in 2- and 4-month-old infants. *Music Percept* 20:187–194.
- Västfjäll D (2002) Emotion induction through music: a review of the musical mood induction procedure. *Musicae Scientiae (Special issue 2001–2002)*:173–211.

- Vieillard S, Peretz I, Gosselin N, Khalfa S, Gagnon L, Bouchard B (2008) Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cogn Emotion* 22:720–752.
- Vrana S (1995) Emotional modulation of skin conductance and eyeblink responses to a startle probe. *Psychophysiology* 32:351–357.
- Watson D, Clark LA (1991) Self-versus peer ratings of specific emotional traits: evidence of convergent and discriminant validity. *J Pers Soc Psychol* 60:927–940.
- Wedin L (1972) A multidimensional study of perceptual-emotional qualities in music. *Scand J Psychol* 13:241–257.
- Witvliet CVO, Vrana SR (2007) Play it again Sam: repeated exposure to emotionally evocative music polarises liking and smiling responses and influences other affective reports facial EMG and heart rate. *Cogn Emotion* 21:3–25.
- Zajonc RB (1968) Attitudinal effects of mere exposure. *J Pers Soc Psychol Monogr* 9(2 Pt 2):1–27.
- Zajonc RC, Shaver P, Tavis C, van Kreveld D (1972) Exposure satiation and stimulus discriminability. *J Pers Soc Psychol* 21:270–280.
- Zentner MR, Kagan J (1996) Perception of music by infants. *Nature* 383:29.
- Zentner MR, Kagan J (1998) Infants' perception of consonance and dissonance in music. *Infant Behav Dev* 21:483–492.
- Zentner MR, Grandjean D, Scherer KR (2008) Emotions evoked by the sound of music: characterization classification and measurement. *Emotion* 8:494–521.

Chapter 6

Tempo and Rhythm

J. Devin McAuley

6.1 Introduction

It is a remarkable feat that listeners develop stable representations for auditory events, given the varied, and often ambiguous, temporal patterning of acoustic energy received by the ears. The focus of this chapter is on empirical and theoretical approaches to tempo and rhythm, two aspects of the temporal patterning of sound that are fundamental to musical communication.

The chapter is organized into five sections. Section 6.2 introduces basic concepts in research on tempo and rhythm and previews the topic areas that are covered in the chapter. Section 6.3 provides a general overview of two contrasting theoretical approaches that have broadly influenced research on tempo and rhythm. Sections 6.4 and 6.5 provide a selective review of research on tempo and rhythm, respectively. The chapter concludes with a summary of key points and a short discussion of promising avenues for future research.

6.2 Basic Concepts

There are two aims of this section. First, the terms tempo, rhythm, grouping, beat, and meter are defined in turn, with an emphasis on how these terms are used in the field of music perception and cognition. Second, a broad overview of the to-be-covered topic areas in research on musical tempo and rhythm is provided.

J.D. McAuley (✉)

Department of Psychology, Michigan State University, East Lansing, MI 48823, USA
e-mail: dmcauley@msu.edu

6.2.1 *Tempo*

In a general sense, tempo simply means the rate (or pace) of events in the environment. A fast (or slow) tempo is a fast (or slow) event rate. In a musical sense, tempo communicates the pace of a piece of music (i.e., how fast or slow it is) and is typically associated with the rate of periodic events (beats) that listeners perceive to occur at regular (equal) temporal intervals. In a notated musical score, the intended tempo is given in terms of beats per minute (e.g., 120 bpm). The beat-per-minute convention is used in research on musical tempo, but tempo is also expressed in the literature as the time interval between successive beats (the beat period). The latter convention has been valuable because it permits a more direct comparison with the broader time perception literature.

Four areas of tempo research are discussed in the chapter: the concept of preferred tempi and limits on the range of time intervals that convey tempo information (Sect. 6.4.1), absolute memory for tempo (Sect. 6.4.2), tempo discrimination (Sect. 6.4.3), and tempo production (Sect. 6.4.4).

6.2.2 *Rhythm*

In music, the term *rhythm* has been used in at least two ways. It can refer either to the sound pattern or to the perception of that pattern. With respect to the sound pattern, rhythm is the serial pattern of durations marked by a series of events; in the case of music, the rhythm of a melody is the serial pattern of durations marked by sounds (notes) and silences (rests). Musical notation specifies this serial duration pattern using relative, rather than absolute, units to represent the duration of notes and rests. With respect to perception, rhythm refers to the perceived temporal organization of the physical sound pattern (i.e., the series of notes and rests). Also commonly associated with the perception of rhythm is a feeling of movement in time (Fraisse 1963; Lerdahl and Jackendoff 1983). Three fundamental characteristics of the perceived temporal organization of music are linked to the concepts of grouping, beat, and meter. Questions regarding grouping concern the *figural coding* of rhythms, while those concerning beat and meter concern the *metric coding* of rhythms (Bamberger 1980; Smith et al. 1994).

Work on rhythm is discussed in three sections: perception of grouping (Sect. 6.5.1), perception of beat and meter (Sect. 6.5.2), and models of rhythm (Sect. 6.5.3). The remainder of this section focuses on definitions for the terms grouping, beat, and meter and how these terms have been used in the literature.

6.2.2.1 *Grouping*

Grouping refers to how a series of notes are perceived to be clustered or grouped together. Research on principles of grouping and their role in the figural coding of

rhythms has a long history, sharing similarities with work on the development of Gestalt principles of perceptual organization (Wundt 1874; Bolton 1894; Woodrow 1909; Wallin 1911). The section on the perception of grouping (Sect. 6.5.1) considers how sound characteristics, such as duration, frequency, and intensity, influence the way listeners perceive the grouping of notes in musical patterns as well as the tendency for listeners to impose grouping structure in the absence of any acoustic cues in the signal.

6.2.2.2 Beat and Meter

A basic response to music is to clap, tap, or move the body in time in a periodic fashion with a perceived pulse or beat (Parncutt 1994; Snyder and Krumhansl 2001; Temperley 2001; Large and Palmer 2002). Beats, by definition, occur at periodic intervals; in a musical score, the beat is notated. However, what is meant by a *perceived beat* in the preceding example is a series of approximately periodic time points in music that stand out in some way to the listener (i.e., they are accented). Beats (either notated or perceived) often coincide with onsets of musical notes (sounded events), but they do not have to. That is, beats can occur on silent musical elements. Ideally, the perceived beat coincides with the beat that a performer intends and/or is notated in the musical score, but whether this is true in practice depends on a variety of factors. Most music evokes a sense of beat, but there are examples of musical styles, such as Gregorian chant, wherein a periodic beat is not readily discernible (Crocker 2000).

Meter refers to the temporal organization of beats on multiple time scales. These time scales form a *metric hierarchy*, such that the beats at each level of the hierarchy periodically coincide. Beats at lower levels of the hierarchy are at a faster tempo than beats at higher levels of the hierarchy. In a musical score, the meter of the music is typically marked by a time signature that specifies two levels of the hierarchy: the primary beat (also called *tactus*) level and one higher level, called the *measure*. Each measure in music is subdivided into an equal number of beats, where each beat has a prescribed (notated) duration (also referred to as the *beat period*). Figure 6.1 provides two examples. In the left half of the figure, a 2/4 time signature specifies a measure that is equally subdivided into two equal-duration quarter notes where each quarter note is assigned a beat; this means that in a 2/4 time signature, there is a 2:1 ratio between the time span of the measure and the primary beat period; this is also called a *duple meter*. In the right half of the figure, a 3/4 time signature, in contrast, the measure is subdivided into three equal-duration quarter notes. In this case, there is a 3:1 ratio between the time span of the measure and the primary beat period; this is also called a *triple meter*. Duple and triple meters are common in Western music. *Meter perception* refers to hearing beats on multiple time scales with some beats heard as more accented (stronger) than others based on the metric hierarchy implied by the temporal structure of the rhythm. Figure 6.1b illustrates metric hierarchies for the two notated musical examples in Fig. 6.1a. Ideally, there is a correspondence between the perceived meter and the meter that a performer intends and/or is notated in the musical score. If the meter

stopwatch or hourglass (Gibbon et al. 1984; Ivry and Hazeltine 1995; Mattel and Meck 2000). A contrasting view is that the mind's clock resembles in many ways a self-sustaining oscillatory process (Large and Jones 1999; McAuley and Jones 2003). Hourglass and oscillator conceptions of the clock have been linked to two general theoretical approaches, commonly referred to in the literature as interval theories and entrainment theories, respectively.

6.3.1 *Interval Theories*

Interval theories are based on an information-processing framework. Formal models developed within this framework typically posit distinct clock, memory, and decision stages of temporal processing (Gibbon et al. 1984; Meck 2003). Interval models have been applied to both perceptual and motor aspects of responding to musical events (Keele et al. 1989; Ivry and Hazeltine 1995; Meck 2003). Perhaps the most influential and tested of these models is scalar expectancy theory (Gibbon 1977; Gibbon et al. 1984; Church 2003). Complementary to scalar expectancy theory (SET) is the Wing and Kristofferson (W&K) model (Wing and Kristofferson 1973), which has been widely applied to rhythmic tapping behavior in an effort to distinguish clock and motor sources of performance variability.

In SET and related interval models, the clock stage involves a pacemaker, which emits over time a continuous stream of pulses that flow into an accumulator via an attention-controlled switch. At the start of a to-be-estimated (target) time interval, the switch closes, allowing pulses to flow into an accumulator; at the end of the to-be-estimated time interval, the switch opens, stopping the flow of pulses into the accumulator. The number of pulses accumulated during the target time interval provides a representation of duration. With each time interval, the accumulator is cleared with the closing of the switch; thus, the switch acts like an arbitrary reset signal, similar to how a stop watch or hourglass is reset. At the memory stage, each estimate of duration is stored in a long-term reference memory; over time, a distribution of duration codes develops. At the decision stage, the current accumulator account is compared with a temporal criterion (a stored duration code) sampled from reference memory to permit a temporal judgment, such as "shorter" or "longer."

The following example illustrates how the general interval approach to timing works. Imagine approaching a stop light in a car for the first time. As the stop light turns red, the switch closes and pulses begin to collect in the accumulator. Similarly, when the stop light turns green the switch opens and the accumulation process stops. The number (count) of the pulses over the temporal extent of the red light provides a representation of the duration of the stop light (a duration code) that can be stored in memory. With many visits to the same stop light, a memory distribution of duration codes develops. Thus, each new visit to the light refines temporal expectations about when the stop light will turn green. Expectations come into play at the decision stage. At the decision stage, the current accumulator count is compared in

a continuous fashion with a sampled count from memory. The sampled count functions as a temporal criterion. When the current count reaches the temporal criterion, then, in essence, “time is up” and the currently timed interval is expected to end. If the objective time interval ends before the count reaches the temporal criterion, then the interval is shorter than expected; conversely, if the count reaches the temporal criterion before the objective time interval ends, then the interval is longer than expected.

Interval theories have generally provided an elegant explanation for a number of timing phenomena (Meck 1996, 2005). With respect to musical tempo and rhythm, interval theories have been applied primarily to tempo discrimination (Sect. 6.4.3). However, interval theories have also been extended to address motor aspects of responding to musical events (Sect. 6.4.4) and to a lesser extent beat and meter perception (Sect. 6.5.3).

6.3.2 *Entrainment Theories*

Entrainment theories derive from a dynamical systems perspective, rather than an information-processing framework. In broad terms, entrainment approaches propose that the tempo and rhythm of everyday events engage people on a moment-to-moment basis through attentional synchrony (Jones 1976; Jones and Boltz 1989; Large and Jones 1999; McAuley and Jones 2003). This approach builds on the observation that entrainment is a widespread phenomenon in nature, with entrainment of human circadian rhythms (e.g., sleep–wake cycle) with various environmental rhythms as one familiar biological example (Moore-Ede et al. 1982). In the case of circadian rhythms, the *driving* rhythm is an environmental rhythm, such as the daily light–dark cycle and the *driven* rhythm is the self-sustaining oscillation generated by pacemaker neurons in the suprachiasmatic nucleus.

Entrainment theories on a millisecond-to-second time scale are similarly based on the concept of self-sustaining entrainable oscillation (Jones and Boltz 1989; McAuley 1995; McAuley and Kidd 1998; Large and Jones 1999; Eck 2002; McAuley and Jones 2003). Applied to music, the driving rhythm is the external musical rhythm and the driven rhythm is a self-sustaining neural oscillation. Unlike with circadian rhythms, the peaks in oscillator amplitude in models of musical entrainment are assumed to represent periodic changes in gross neural activity, rather than the response of individual pacemaker cells, with the period of the oscillation providing a potential referent for judgments about relative timing; for the express purposes of chapter goals, the amplitude peaks is referred to as “beats.” Note that these “beats” of the entrainment model are subjective beats that may or may not correspond precisely with objective beats in the signal. That is, musical events (notes) may be either temporally aligned (in phase) or misaligned (out of phase) with the periodic timing of peaks in oscillation amplitude (subjective beats). Rather than distinguishing between clock and memory stages of temporal processing, the entrainment approach assumes that individuals make judgments about the timing of

a sequence of musical events by detecting the synchrony/asynchrony of successive stimuli with these periodic beats. Thus, entrainment models forgo an explicit memory comparison between two coded durations in favor of dynamic information afforded by synchrony versus asynchrony between the internal driven rhythm and the external driving rhythm.

Musical events that arrive unexpectedly “early” (i.e., before an amplitude peak/beat) provide evidence that the local tempo of a sequence of musical events is accelerating, while musical events that arrive unexpectedly “late” (i.e., after a beat) imply that the local tempo of a sequence of musical events is “decelerating.” McAuley and Jones (2003; see also Jones and Boltz 1989) refer to the discrepancy between expected and actual onset times as *temporal contrast*. Critically, two types of adaptive processes are assumed to operate on temporal contrast to facilitate entrainment: phase correction and period correction. The role of phase correction is to align amplitude peaks (expected time points) with musical event onsets. The role of period correction is to adjust the period of the internal oscillator so that it eventually matches salient time intervals marked by the sequence of musical events. The range of rates (tempi) that afford stable entrainment (synchronization) is referred to as an *entrainment region*. Dynamic attending theory (DAT) is a generalization of entrainment theory, whereby the internal driven rhythm is conceptualized as an attentional rhythm (Jones 1976; Large and Jones 1999).

Empirical support for entrainment theories and more broadly DAT comes from a range of sources. First, studies of overt motor tracking of event sequences indicate less variability and greater accuracy in responding to rhythmically simple events than to complex events (Jones and Pfordresher 1997; Large et al. 2002; Large and Palmer 2002; Pfordresher 2003). Second, in perceptual monitoring tasks, simple rhythms enhance performance with rhythmically expected targets, suggesting that attentional synchrony has a facilitating influence on detection of pitch, timbre, or time changes (Jones et al. 1982, 2002; Klein and Jones 1996; Jones and Yee 1997; Barnes and Jones 2000; McAuley and Jones 2003). Third, rhythms facilitate detection of temporal order of target pitches that are embedded in longer sequences (Jones et al. 1981). Fourth, in recognition memory tasks people mistakenly identify decoy melodies as target melodies when the decoys occur in a target’s rhythm (Jones and Ralston 1991). Finally, the recall of pitch sequences is better when various accents (e.g., pitch skips, contour changes) are timed regularly rather than irregularly (Boltz and Jones 1986).

In sum, there are a number of key differences between entrainment and interval theories. First, from an interval perspective, duration is explicitly represented as a duration code that is stored in memory, whereas from an entrainment perspective, duration is implicitly represented by the oscillator period. Second, assumptions about responses to stimulus onsets differ. Interval models assume arbitrary reset of the pacemaker-accumulator clock with each stimulus onset, whereas entrainment models assume more gradual correction of oscillator phase and period. Third, with interval models, successive time estimates are independent, whereas in entrainment models these are dependent. Finally, with respect to duration/timing judgments,

interval models involve explicit comparison of two stored duration codes whereas entrainment models involve a phase-based temporal contrast metric.

Having distinguished two major theoretical perspectives on tempo and rhythm and covered basic terminology, the next section turns to a synthesis of past and current approaches to the study of musical tempo and the relationship of this work more broadly to research on time perception.

6.4 Tempo

What is the function of the tempo of a piece of music? For one thing, tempo communicates emotion, with fast music tending to be perceived as “happy” and slow music tending to be perceived as “sad” by both children and adults (Dalla Bella et al. 2001; Gagnon and Peretz 2003; see Schellenberg, Chap. 5). More generally, tempo helps listeners track musical events as they unfold in time and enables predictions about when future events are likely to occur. This latter function of tempo appeals beyond the study of music because of the importance of prediction to general cognitive processes.

6.4.1 *Tempo Limits and the Concept of Preferred Tempo*

The range of tempi over which beats are perceived is limited (Fraisse 1963, 1982). If music is performed too quickly, successive sounds become indistinguishable. Conversely, if music is performed too slowly, rhythmic organization tends to fall apart, leaving only a series of isolated sounds. Between the two extremes, music and other sound patterns have perceivable rhythm. A conservative estimate for the upper (fast) tempo limit is around 100 ms between sounds (Friberg and Sundström 2002; London 2004), while an estimate for the lower (slow) tempo limit is around 2.5 s between sounds; reported values for fast and slow tempo limits vary considerably in the literature, however (Fraisse 1982; Clarke 1999; McAuley et al. 2006). Fraisse (1963) referred to the range of time intervals between 0.1 and 2.5 s as the “psychological present”; see also James (1890). Pöppel and colleagues link the slow tempo limit to the temporal capacity of working memory (Szlag et al. 1996; Pöppel 1997).

Other work has considered motor tempo limits on people’s ability to clap, tap, or generally move in synchrony with music. Notably, motor tempo limits closely parallel those observed for perception. The upper (fast) tempo limit for 1:1 synchronization (e.g., one tap per beat) tends to be partly constrained by the rate at which an individual can move, but is also likely driven by the increased proportional variability that is found at fast rates. For hand tapping, the upper (fast) limit is approximately 150–200 ms between taps (Fraisse 1982; Repp 2003; McAuley et al. 2006) and the lower (slow) tempo limit is about 2 s between taps. For time intervals longer

than 2 s or so between taps, participants tend to have difficulty predicting when the musical event (e.g., tone) will occur and simply reacting to, rather than anticipating, tone onsets. Research on tempo limits associated with synchronizing movements with music is generally consistent with entrainment theories, which have proposed that synchronization should be most accurate within a limited entrainment region, with less stable synchronization performance and increased variability outside that region. At least one developmental study has suggested that the entrainment region is narrower for children and older adults compared with young adults (McAuley et al. 2006).

Within the tempo limits that define perceivable rhythms and afford synchronization, individuals demonstrate clear tempo preferences. The concept of a preferred tempo has been widely studied (Stern 1900; Wallin 1911; Fraisse 1963, 1982; Jones 1976; McAuley et al. 2006) and names given for preferred tempo in the literature vary considerably; these include “mental tempo,” “personal tempo,” “psychic tempo,” and “internal tempo” (Stern 1900; Rimoldi 1951; Mishima 1956; Fraisse 1963, 1982; Boltz 1994; Vanneste et al. 2001). The various names for preferred tempo reflect different assumptions about what preferred tempo means. Does preferred tempo simply index a preferred rate of spontaneous motor activity, or a preferred rate of listening, or does it measure a broader cognitive tempo preference? From an entrainment perspective, the concept of a preferred tempo is linked to the intrinsic period of the driven oscillator (McAuley et al. 2006). From an interval perspective, preferred tempo is sometimes associated with clock speed (i.e., rate of pulse accumulation; Vanneste et al. 2001). Independent of theoretical orientation or assumptions about the underlying meaning of preferred tempo, assessments of preferred tempo have generally emphasized either spontaneous motor measures or perceptual measures. Work on motor and perceptual measures of preferred tempo is reviewed in turn next.

6.4.1.1 Spontaneous Motor Tempo

Stern (1900) was one of the first researchers to suggest that the tempo of spontaneous motor activity provides some insight about the pace of mental activity. To measure this, Stern asked individuals to tap their hands on a table at a rate they considered just right (not too fast or too slow). This assessment of the tempo of spontaneous motor activity has become one of the most widely used measures of preferred tempo. Not all researchers agree, however, that spontaneous motor tempo reflects the pace of mental activity. Nonetheless, a common assumption of much work in this area is that the preferred tempo of spontaneous rhythmic motor activities, such as walking or clapping, does provide some insight about the pace of an internal “mental” clock involved in the perception of time (Boltz 1994; Vanneste et al. 2001).

The most representative value of spontaneous motor tempo (SMT) reported in the literature for adults is around 600 ms (Fraisse 1982). Across studies a representative range extends from approximately 300 ms to 800 ms (Frischeisen-Köhler 1933;

Mishima 1956; Smoll and Schutz 1978; Fraisse 1982; McAuley et al. 2006). One notable exception is Collyer et al. (1994), who report a bimodal distribution of spontaneous motor tempi with modes around 272 ms and 450 ms.

Although 600 ms is a representative value of SMT, there are also large individual differences. Observed values in individual assessments of SMT vary widely; SMT can be as short as 200 ms or as long as 1,600 ms (Drake et al. 2000; McAuley et al. 2006). Despite large individual differences, measures of spontaneous motor tempo tend to be reliable. Spontaneous motor tempo is very stable within a given production, with the sequence of produced intervals varying on average by $\approx 5\%$. This is similar to the degree of variability observed in assessments of duration discrimination. When multiple measures of SMT are taken, correlations across these measures range from 0.75 to 0.95 (Harrel 1937; Rimoldi 1951; McAuley et al. 2006).

6.4.1.2 Preferred Perceptual Tempo

Whereas spontaneous motor tempo refers to the natural or preferred rate of rhythmic motor activity (e.g., tapping), preferred perceptual tempo refers to the rate of a series of sounds or lights that is judged to be neither too fast, nor too slow, but appears to be “just right” (Fraisse 1982; McAuley et al. 2006). Early work on preferred perceptual tempo (PPT) concerned identifying an *indifference interval* that individuals perceived as neither too short nor too long (Vierordt 1868; Woodrow 1951). Like SMT, the most commonly reported value for this interval is around 600 ms (Frischeisen-Köhler 1933; Mishima 1956; McAuley et al. 2006), but a wide range of values have also been reported over the years (Wallin 1911; Woodrow 1951; Fraisse 1982; McAuley et al. 2006). Thus, notably, SMT and PPT have comparable frequencies. However, evidence for the correlation between the two is mixed (Fraisse 1982). Some of the strongest support for a correlation between the two is reported by McAuley et al. (2006), who found a large, positive, correlation between SMT and PPT, i.e., near 0.75. Such correlations support the view that motor and perceptual tempo preferences have a common psychological basis.

6.4.1.3 Factors Affecting Preferred Tempo

Developmental studies of preferred tempo reveal that both preferred perceptual tempo and spontaneous motor tempo slow with increased age (Drake et al. 2000; Vanneste et al. 2001; McAuley et al. 2006). In the most extensive of these studies, McAuley et al. (2006) examined PPT and SMT for participants between the ages of 4 and 95 years and found that the preference for slower sequences (PPT) increased systematically with age, thus paralleling a similar age-related motor slowing in SMT. Moreover, PPT was also found to highly correlate with SMT across all age ranges. For children between the ages of 4 and 7, preferred tempo was typically between 300 and 400 ms; for adults, preferred tempo was around 600 ms, while for older adults preferred tempo was close to 700 ms. Drake et al. (2000) and Vanneste

et al. (2001) reported similar age-related trends. In sum, there is emerging support for the age-related slowing of perceptual and motor measures of preferred tempo across the lifespan. Interpreted within an interval model framework, these data support the view that the internal clock slows with increased age. From an entrainment model perspective, these data suggest an age-related lengthening of intrinsic oscillator period.

Fraisse (1982) reported that differences in SMT between two identical twins are no different than two productions of SMT by the same subject, but that differences in SMT between two fraternal twins are as large as those found between two individuals selected at random. This suggests that there may be a genetic basis for preferred tempo, but there is limited evidence to support this view. For example, it appears that elements of preferred tempo are learned. Drake et al. (2000) showed that preferred tempi of children with musical training tend to shift to adult levels sooner than children without musical training. Thus, some of the observed developmental changes in preferred tempo may be a consequence of experience, rather than maturation, with more musical experience speeding the rate of developmental change. A learning perspective is consistent with other work showing that even on the time scale of an experiment, participants develop a general sense of the average pace of the events they experience, and this average pace can serve as a referent for participants judgments about the tempo (or duration) of events in their environment (Jones and McAuley 2005; Miller and McAuley 2005).

Aside from effects of age, individual differences in preferred tempo do not appear to be associated with other intrinsic factors, such as gender, handedness, or body size. There is also little evidence to link preferred tempo to specific physiological variables, such as heart rate. One study suggestive of a link between preferred tempo and physiological variables is the work of Boltz (1994), who showed that exposure to annoying sounds, such as a car horn, hypothesized to increase general arousal levels, tended to speed up preferred tempo, while exposure to relaxing music, hypothesized to decrease general arousal levels, tended to slow down preferred tempo. Boltz further showed that changes in preferred tempo predictably influenced individuals' judgments about learned event durations. This latter finding supports a link between preferred tempo and the speed of an internal clock if a theorist is interval model inclined, or a link to intrinsic oscillator period if one is more entrainment oriented.

6.4.2 *Absolute Memory for Tempo*

Music typically can be performed at a range of tempi and still retain its identity (Andrews et al. 1998). For example, "Happy Birthday" can be sung fast or slow and still be recognized by listeners as "Happy Birthday." This flexibility in recognition occurs because the identity of a piece of a melody does not, typically, depend on the absolute durations of individual notes, but rather on the pattern of relative durations; indeed, this aspect of music perception is captured by musical notation, which

specifies only in relative terms how long each note should be sung or played. The fact that tune identity typically depends on the relative, rather than absolute, duration of notes, raises the interesting question of whether individuals possess memories for the absolute (e.g., millisecond) temporal features of music even though it is the relative patterning of durations that matters most for recognition.

Researchers interested in this question have tended to focus on whether people possess absolute memory for tempo (Farnsworth et al. 1934; Halpern 1988; Bergeson and Trehub 2002; Levitin and Cook 1996). Bergeson and Trehub (2002) compared the tempi of mothers' singing to infants to the tempi of their spoken utterances and found that for singing, repeated productions of the same song demonstrated remarkable stability in tempo, as well as pitch and rhythm. With respect to tempo, differences across song productions varied by only around 3%. For speech, in contrast, repeated productions of the same utterance showed much less stability in pitch and tempo, although speech rhythm tended to be preserved. Tempo differences for repeated productions of the same utterance were on the order of 20%.

Other work suggests that whether listeners retain the absolute temporal features of music may depend on the degree to which they have directly experienced tempo variations in a piece of music. Music with little variation in the experienced tempo, such as commercial recordings of pop songs heard on the radio, are said to have a tempo standard; that is, repeated exposures to the song are always at about the same tempo. Levitin and Cook (1996) examined absolute memory for the tempo by having adult listeners without musical training sing pop songs from memory using a cued list of songs typically heard on the radio. Consistent with the view that individuals encode the absolute temporal features of music, song productions of 72% of the participants were within 8% of the actual tempo of the commercial recordings frequently heard on the radio. Notably, this degree of variability was no different than the typical range of tempo discrimination thresholds reported in the literature. Moreover, the correlation between produced and actual tempo was 0.95. In contrast, vocal productions of a set of familiar folk songs without a tempo standard revealed substantially more variability. Moreover, because people were shown to be capable of producing song tempi at very fast and very slow rates, this ruled out the possibility that the precise tempo memory observed for pop songs was simply due to production (performance) constraints.

Halpern (1988) considered how closely the remembered (imagined) tempo of familiar music matched the perceived tempo using a set of familiar songs that notably did not have a tempo standard. For the perception version of this task, participants listened to a song and adjusted a metronome until they arrived at their preferred tempo for the song. For the imagery version, participants imagined the tune they heard and adjusted clicks of a metronome until they matched beats of the imagined song. Providing support for the view that tempo is represented in auditory imagery, the correlation between perceived and imagined song tempi was around 0.6. Moreover, there was a tendency for the imagined tempo of a song to be faster than the perceived tempo when the perceived tempo was slow, whereas the reverse was true when the perceived tempo was fast. This suggests that the imagined tempo of a song tends to gravitate to a mean rate. Notably, the average time interval

between beats for perceived and imagined tempi was around 600 ms, which is consistent with work on preferred tempo. Taken together, representation of absolute tempo information in auditory imagery appears to include both a local learned component that reflects the tempo of the particular piece of music and a more global temporal context component. It is not clear whether the more global component simply reflects the average tempo experienced by listeners or an individual's intrinsic preferred tempo, which conceivably may have a genetic basis.

In sum, there is converging evidence that musicians and nonmusicians alike develop fairly precise memories for the absolute tempo of music, as long as the music has a tempo standard (i.e., it is almost always performed at about the same tempo). This is not true for music without a tempo standard or for the tempo of spoken utterances. Systematic deviations in memory for the absolute tempo of songs with a tempo standard and for speech are reminiscent of some of the work on preferred tempo. Having discussed tempo preferences and listeners' ability to develop long-term memory for tempi, the next section addresses listeners' ability to detect changes in tempo.

6.4.3 *Tempo Discrimination*

Another area of tempo research addresses listeners' ability to detect changes in tempo. Two questions have guided this research. The first concerns whether tempo discrimination thresholds obey Weber's law; the second question concerns the impact of multiple sequence intervals on thresholds.

6.4.3.1 **Weber's Law and the Multiple-Interval Advantage**

Studies of tempo discrimination have involved both auditory and visual sequences (Michon 1964; Schulze 1978, 1989; Drake and Botte 1993; ten Hoopen et al. 1994; Ivry and Hazeltine 1995; Vos et al. 1997; McAuley and Kidd 1998; Grondin 2001a). A central debate in this literature concerns whether discrimination thresholds obey Weber's law. Weber's law is assessed by either perceptual measures or motor measures. Motor assessments of Weber's law is discussed in Sect. 6.4.4. With respect to perceptual measures, if Weber's law holds for a range of tempi (time intervals), then the just-noticeable difference in tempo (ΔT) between two sequences should be a constant proportion (or percentage) of a base time interval, T , where the value of T is typically given by a fixed referent time interval marked by successive tone onsets in one of the two sequences. A formal way of expressing Weber's law is that $\Delta T/T$ should be equal to a constant, k . The value of k is called a Weber fraction. A Weber fraction of 0.1 indicates that a listener can detect a 50-ms difference given a T of 500 ms, but would need at least a 100-ms difference for a T of 1,000 ms. Empirical studies have examined the simple version of Weber's law described in the preceding text and generalized variants; see Grondin (2001b) for a comprehensive review.

Many of the studies addressing the issue of Weber's law in a musical context have focused on tempo judgments about isochronous (equal interval) rhythms. Figure 6.2 illustrates a typical tempo discrimination task, in which listeners hear two monotone sequences, a standard sequence followed by a comparison sequence, and they must judge the tempo of the comparison relative to the standard. In general, tempo discrimination is poorer when the two sequences are irregularly timed than when they are regularly timed sequences (Drake and Botte 1993).

Michon (1964) measured listeners' ability to detect tempo differences for inter-onset intervals (IOIs) between 67 and 2,700 ms using relatively long sequences and very well practiced subjects. Counter to Weber's law, he found a minimum relative just-noticeable difference (JND) that was as small as 1.0% for a 100-ms IOI with a secondary minimum region around 2.0% for IOIs between 300 and 1,000 ms. Both minimums were much lower than the approximate 6% relative JNDs reported for isolated intervals. This latter finding suggested an advantage afforded by multiple sequence intervals.

The issue of single versus multiple intervals was considered in detail by Drake and Botte (1993). They examined listeners' ability to detect tempo differences in isochronous sequence for IOIs between 100 and 1,500 ms for 1, 2, 4, and 6 interval sequences. For single-interval sequences, tempo relative JNDs were approximately 6%, which is a similar value to that found in duration discrimination studies (Woodrow 1951; Creelman 1962; Small and Campbell 1962; Abel 1972; Getty 1975; Allan 1979; Drake and Botte 1993). For multiple-interval sequences, thresholds improved, on average, to 3%. Best performance was found for 6-interval sequences for a 400-ms tempo with the reported threshold slightly below 2%. Similar to Michon (1964), relative JNDs for tempi between 100 and 1,500 ms were not constant, as predicted by Weber's law, but rather were a U-shaped function. This finding is consistent with the view taken by some that tempo discrimination is best (i.e., thresholds lowest) at a listener's preferred tempo. Relative JNDs were approximately constant for IOIs between 300 and 800 ms, increasing for both shorter and longer values.

Overall, Drake and Botte observed more improvement with increasing number of intervals at fast tempi than at slow tempi, which is also in line with the work of Michon (1964). Results consistent with these findings were also reported by McAuley and Kidd (1998) for time intervals between 100 and 1,000 ms for 1- and 3-interval sequences; increasing the number of sequence intervals reduced thresholds, especially at the faster tempi.

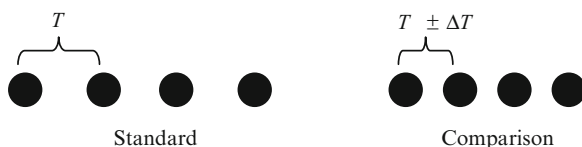


Fig. 6.2 Illustration of a tempo-discrimination task. Two monotone sequences, a four-tone standard sequence with a fixed time interval, T , between tone onsets, followed by a four-tone comparison sequence with a fixed time interval $T \pm \Delta T$. The value of ΔT typically varies from trial to trial. The listeners' task is to judge the tempo of the comparison sequence relative to the standard

In sum, researchers considering the question of Weber's law for tempo discrimination and effects of sequence length (number of equal intervals) on thresholds have found that Weber's law holds only within a limited range of tempi. Notably, this optimal tempo zone encompasses commonly reported values for preferred tempo. Moreover, as the number of equal intervals increases, listeners' ability to detect changes in tempo improves. Across studies JNDs for single-interval sequences are typically on the order of 6%, while JNDs for multiple-interval sequences are sometimes less than about 2% (Michon 1964; Drake and Botte 1993; Friberg and Sundberg 1995).

6.4.3.2 Locus of the Multiple-Interval Advantage

Within an interval-model framework, reduced tempo sensitivity (increased threshold) is typically attributed to increased variability of clock, memory, or decision processes. From this perspective, Drake and Botte proposed to explain improvements in tempo discrimination thresholds associated with the number of sequence intervals using a multiple-look model whereby each equal time interval in an isochronous standard sequence provides an independent but variable estimate of sequence tempo. They hypothesized, as have others, that listening to the standard sequence leads to a series of independently sampled estimates of the tempo of the standard sequence, which are averaged to form an aggregate memory trace (Keele et al. 1989; Schulze 1989; Drake and Botte 1993; Ivry and Hazeltine 1995). As the number of independent "looks" increases, the average sampling error between the estimated and actual standard tempo decreases, leading to lower discrimination thresholds. Drake and Botte predicted that the JND in tempo, taken as the standard deviation of the sampling distribution, should decrease inversely to the square root of the number of standard sequence intervals, as shown here:

$$JND_n = \frac{JND_1}{\sqrt{n}} \quad (6.1)$$

In this model, JND_1 is the observed JND for a single-interval standard sequence and JND_n is the predicted JND for an n -interval standard sequence.

Studies have reported data consistent with the multiple-look model for both auditory and visual sequences for tasks involving time-interval perception as well as production (ten Hoopen and Akerboom 1983; Ivry and Hazeltine 1995; Rousseau and Rousseau 1996; McAuley and Kidd 1998; McAuley and Jones 2003). There are notable exceptions, however. Some studies have reported mixed results (Schulze 1989; Hirsh et al. 1990; Grondin 2001a), whereas others have found no multiple-interval advantage (ten Hoopen et al. 1994; Pashler 2001). One factor preventing a clear interpretation of some of this research is that the numbers of standard and comparison intervals have sometimes covaried, making the locus of the multiple-interval advantage unclear (Drake and Botte 1993; Grondin 2001a). That is, does the multiple-interval advantage occur because of multiple intervals in the first (standard) sequence, the second (comparison) sequence, or both?

Miller and McAuley (2005) investigated the locus of the multiple-interval advantage by independently varying the number of standard and comparison intervals in standard-comparison pairs of isochronous tone sequences. They found, somewhat surprisingly, that with a *fixed* standard tempo on each trial, the multiple-interval advantage occurs because of multiple intervals in the comparison sequence, rather than multiple intervals in the standard sequence. However, when tempo of a standard sequence varies from trial to trial, both the standard sequence and the comparison sequence contribute to the multiple-interval advantage; see also Grondin and McAuley (2009).

To account for distinct contributions of the number of standard and comparison intervals to tempo discrimination thresholds, Miller and McAuley (2005) proposed a generalized multiple-look (GML) model, which extended the original multiple-look model of Drake and Botte (1993) to measure average sampling error for two independent samples corresponding to the number of sampled intervals from the standard and comparison sequences, respectively. As with the Drake and Botte multiple-look model, thresholds in the GML model are predicted to be inversely related to the number of equal sequence intervals. However, unlike the Drake and Botte model, a weight parameter, w , in the GML model permits the threshold contribution of the standard and comparison sequences to vary. One general observation about multiple-look models and related interval approaches is that although they provide a descriptive account of tempo threshold data, a common weakness is that they often do not make explicit predictions about other dependent measures such as points of subjective equality (see Jones and McAuley 2005).

The entrainment approach offers an alternative account of tempo discrimination data. According to this perspective, an effect of the number of standard intervals on tempo thresholds is due to period correction processes, whereas an effect of the number of comparison intervals on tempo thresholds is due to listeners' reliance on temporal contrast information (i.e., judgments about how aligned successive tone onsets in the comparison sequence are with internally generated beats). Thus, increasing the number of standard intervals should produce improvements in tempo sensitivity in situations that require substantial period correction (e.g., when the standard sequence tempo differs from the average tempo) but not in situations that require little or no period correction (e.g., when the standard sequence tempo is at the average tempo). Conversely, increasing the number of comparison intervals is likely to lead to improvements in tempo sensitivity in situations that require little or no period correction (i.e., when there is a close match between the period of an induced internal oscillation and the standard interval). It is precisely in the latter situations where relative phase discrepancies of tone onsets in the comparison sequence (temporal contrasts) provide the most reliable information about the relative tempo of the comparison sequence ("faster" or "slower"). Empirical data from time judgment tasks have supported these entrainment model predictions (Large and Jones 1999; Barnes and Jones 2000; McAuley and Jones 2003; Miller and McAuley 2005; Jones and McAuley 2005).

6.4.4 Produced Tempo

Some of the same issues addressed in research on tempo discrimination discussed in the preceding section emerge in research on performance aspects of music. As with tempo discrimination, a key question concerns whether the tempi that people produce in music performance or in simply tapping along to music obey Weber's law. This question has been most often addressed in the context of a synchronize-continue tapping task, first introduced by Stevens (1886). In its simplest form, synchronize-continue tapping requires individuals to generate a series of finger taps in synchrony with a metronome set to a particular tempo (or target time interval, T), and then to continue tapping at the same rate in the absence of the pacing stimulus. Variants of this paradigm have asked individuals to tap in synchrony with the beat of a musical excerpt, at subdivisions or multiples of the beat (i.e., at different levels of the metric hierarchy), and have compared tapping to simple and complex meters (Patel et al. 2005; Repp 2005; Snyder et al. 2006). Performance across all versions of the paradigm is typically evaluated by the phase alignment of taps with sounds during the synchronization phase, as well as the mean and variance of both the produced intervals for both the synchronization and continuation phases of the task. For the purposes of comparison to work on perceived tempo, the emphasis of this section is on the simple version of the paradigm involving isochronous sequences and assessments of the mean and variability of produced time intervals (i.e., the produced tempo).

A widely cited modeling approach to synchronize-continue tapping performance is the interval model of Wing and Kristofferson (termed the W&K model; Wing and Kristofferson 1973). In the W&K model, synchronize-continue tapping entails a series of responses (taps) to each in a series of multiple intervals. Each tap is triggered by an internal clock which, in synchronizing to an isochronous sequence, reflects an encoding of T . That is, stimulus tempo, expressed as a time interval T , determines the number of pacemaker pulses corresponding to the stored interval code (C) and is used to meter out each time interval between successive taps. Further, in continuation tapping, where people must sustain the induced sequence rate in the absence of tones, this model predicts the n th produced tapped interval: I_n . Specifically, this interval is given by an additive combination of the n th interval code (C_n) and peripheral (motor) delays associated with taps that initiate (D_{n-1}) and terminate (D_n) this interval:

$$I_n = C_n + D_n - D_{n-1} \quad (6.2)$$

According to this model, the variability of produced intervals (I) derives from two sources: clock variance and motor variance. Moreover, with an assumption of independence between clock and motor components, it is possible to decompose the total tapping variance (σ_I^2) into separate estimates of clock (σ_C^2) and motor (σ_D^2) sources of variability (Wing 1980; Wing and Kristofferson 1973).

As with interval models of tempo discrimination described in the preceding section, the W&K model makes predictions about accuracy and variability of performance

(here motor performance) as a function of tempo, T . With regard to accuracy, a central prediction of the W&K model holds that the averaged produced interval in continuation tapping will approximate T ; no systematic over- or underestimations of T are predicted. This is because clock values (C_n) should always center on the target interval, T . Although this prediction generally holds for young adults, children and older adults frequently fail to maintain a constant rate of tapping (Ivry and Keele 1989; Williams et al. 1992; Greene and Williams 1993; McAuley et al. 2006), suggesting rather periodic drift to a preferred tempo. In these cases, period drift is typically treated as a nuisance variable and eliminated from the time series using a linear or nonlinear detrending procedure (Ogden and Collier 1999). In contrast, for some entrainment models detrending eliminates an important component of behavior. This is because these models assume that an intrinsic oscillator period, which gradually adjusts during synchronization tapping, will exhibit a diagnostic trend of drifting back to an intrinsic preferred rate in continuation tapping.

A second prediction of the W&K model concerns tapping variability. This revives issues surrounding tempo and Weber's law in tempo discrimination. For the W&K model the decomposition of total tapping variance into clock and motor components is crucial because it leads to the prediction that clock variance increases linearly with the target interval, T , whereas motor variance remains constant. This is an important prediction, not only because it is the mainstay of the W&K model, but also because it violates Weber's law. According to Weber's law, the standard deviation of produced intervals should be linearly related to tempo: that is, as T increases, the standard deviation of taps, which reflect a motor JND, should increase proportionally. However, because the W&K model assumes clock variance, σ_c^2 , increases linearly with T , it predicts that observed tapping variance will increase much more with slower tempi than predicted by Weber's law. Although some applications of the W&K model to continuation tapping of young adults have been moderately successful (Wing 1980) in explaining tapping variability, their generality remains in doubt. Such contrasting predictions about tempo and Weber's law remain the source of much lively debate in the literature.

Using the W&K approach, researchers have explored the possibility that children and older adults show increased clock variance, increased motor variance, or both, relative to young adults. From this work, it is clear that tapping of young children is more variable than tapping of young adults for the reported tempi (Ivry and Keele 1989; Greene and Williams 1993; Geuze and Kalverboer 1994). Less clear is the extent of age differences between older and younger adults (Duchek et al. 1994; Vanneste et al. 2001; Krampe et al. 2005). Many of these studies have attributed observed age-related differences in tapping variability to clock rather than motor sources. One limitation of some of the research using the W&K model is that despite theoretical predictions about the relationship between tempo and variability, the W&K model often has been evaluated at only a single tempo. When a broad range of tempi are considered, the W&K model does not generally fare too well. Rather, consistent with the Drake and Botte (1993) research on tempo discrimination and the multiple-interval advantage discussed in the preceding section, Weber's law appears to provide a better account of the data, but only for a

limited range of tempi. That is, for a restricted range of target intervals, T , the standard deviation, not the variance, of produced intervals increased linearly with T , resulting in a constant Weber fraction. Outside this range at both faster and slow tempi, the Weber fraction tends to increase.

Having discussed theoretical approaches to musical tempo and key empirical findings, the next section turns to a synthesis of past and current approaches to the study of musical rhythm.

6.5 Rhythm

Empirical and theoretical approaches to rhythm have often distinguished the perception of grouping and the figural coding of rhythms from the perception of beat and meter and the metric coding of rhythms (Bamberger 1980; Povel and Essens 1985; Ross and Houtsma 1994; Hébert and Cuddy 2002). Perception of grouping involves the segmentation of a rhythm into clusters of elements (e.g., a group of two, following by a group of three). A figural code based on principles of grouping preserves the number of tones in each successive group (figure) and the number of figures in each repetition of the rhythm. Perception of beat and meter involves the perception of temporal regularity on multiple hierarchically related time scales. Metric coding minimally involves the representation of at least two level of the metric hierarchy and preserves the relative durations of elements of the rhythm. All rhythms do not necessarily afford metric coding (i.e., fit within a metric framework; see Povel and Essens 1985). Moreover, rhythm discrimination can rely on differences in grouping structure (figural coding), metrical structure (metric coding), or both (Handel 1998).

A selective review of rhythm research is presented in three sections. Section 6.5.1 describes work on perception of grouping. Section 6.5.2 describes work on perception of beat and meter. Section 6.5.3 considers several influential models of rhythm.

6.5.1 *Perception of Grouping*

A range of acoustic cues, including frequency, duration, and amplitude (intensity) have the potential, depending on their patterning, to convey to the listener a sense of inherent sequence organization or structure (Fraisse 1956; Handel 1989). The focus of this section is on the perception of grouping, namely that some sequence elements belong together (i.e., they are grouped) whereas others do not; moreover, within a group, some elements are accented, while others are not. There is substantial evidence that grouping affords a figural coding of rhythms that is distinct from the coding of metrical structure, and which plays an important role in listener ability to discriminate and remember rhythms (Handel 1998).

Figure 6.3 illustrates how various acoustic cues influence perceived grouping of elements of a rhythm. First, if every second or third element in a sequence is accented by increasing its intensity, then the elements tend to be perceived to be grouped into twos or threes, respectively, with the element of increased intensity beginning the group (Fig. 6.3a); moreover, the time interval between groups tends to be incorrectly perceived as longer than the time intervals separating elements within the group (Bolton 1894; Woodrow 1909).

Second, if the duration of every second or third element in an otherwise isochronous sequence is lengthened, then elements of the sequence are often perceived in

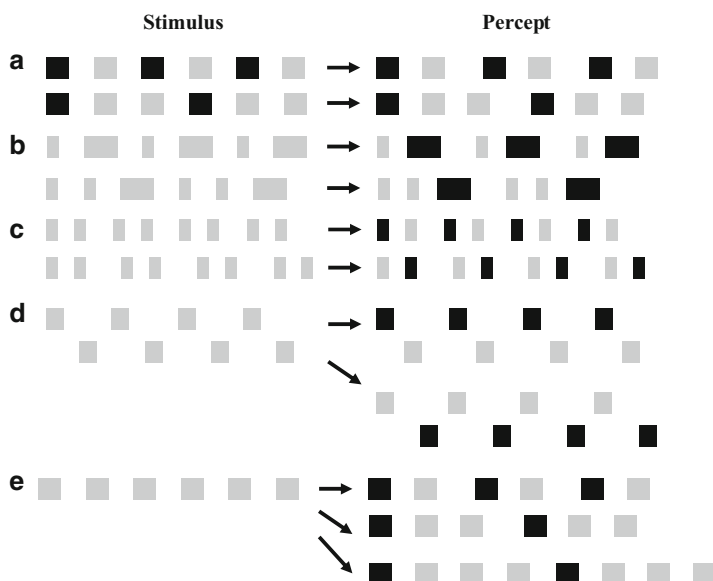


Fig. 6.3 Examples of principles of grouping. The left half of the figure shows a series of sound sequence examples (wider squares indicate lengthening of a musical element, while darker shading indicates an increase in sound intensity). The right half of the figure illustrates typical percepts for each example (darker shading indicates perceived accentuation of the musical element). (a) The intensity of every second or third element in an isochronous sequence is increased, leading to grouping by twos or threes, respectively. The more intense (louder) tone begins each group with the perceived time interval between groups longer than the perceived time intervals separating elements within a group. (b) The duration of every second or third element in an isochronous sequence is lengthened, leading to grouping by twos or threes, respectively. The longer element ends each group with the perceived time interval between groups longer than the perceived time intervals separating elements within a group. (c) Every other time interval between tone onsets is lengthened, leading to binary grouping. For differences between adjacent time intervals that are relatively small, the first element of each group is perceived as accented, whereas for differences that are relatively large, the second tone of each group is perceived as accented. (d) Alternating patterns of high and low tones lead to binary grouping with either the high or low tone perceived as accented and beginning the group. (e) Common subjective rhythms for isochronous sequences of identical elements

groups of two or three, with lengthened elements perceived as accented (Woodrow 1909); here as end accents of each group (Fig. 6.3b). As with intensity accentuation, the time intervals in between groups tend to be subjectively longer than the time intervals separating elements within a group (Handel 1989; Woodrow 1951).

Third, grouping of elements of a rhythm is also affected through variation in the time intervals between element onsets (i.e., holding constant element durations). Povel and Okkerman (1981) showed that when every second or third inter-onset interval is lengthened, then individuals tend to hear the groups of two or threes, respectively. In this case, when the difference between two successive intervals is relatively small, the first element of each group tends to be perceived as accented, whereas when this difference is relatively large, the final element of each group tends to be perceived as accented (Fig. 6.3c); see also the classic work of Garner (1974).

Fourth, frequency (pitch) cues also affect grouping (Fig. 6.3d). For pitch cues, there is a tendency to perceive the rhythmic organization of sequences according to repeated pitch patterning (Woodrow 1911; Steedman 1977). For example, when individuals listen to an isochronous sequence of tones of equal amplitude and duration that alternate between a fixed high and fixed low frequency (e.g., HLHLHLHL), they tend to hear a binary grouping of elements with accents on either the high or low tone, with the accented element beginning the group (Woodrow 1909, 1911).

Finally, some of the earliest work by Bolton (1894) on principles of grouping and the figural coding of rhythms found that even for isochronous sequences of identical sounds, listeners tend to perceive the elements of the sequences to be grouped in twos, threes, or fours, with the first element of each group judged to be more accented than the others (Fig. 6.3e). As the tempo of a rhythm increases, there is a tendency for the number of elements perceived to form a group to increase, suggesting that there may be an intrinsic preferred total duration for each group (Harrel 1937).

Much work has assumed that principles of grouping are universal. However, several studies have shown that performers use their knowledge of musical structure to emphasize the grouping of elements by increasing the intensity or duration of the final element in a group (Drake 1993; Drake and Palmer 1993; Repp et al. 2002). Moreover, at least one recent study has shown that perception of grouping can be dependent on cultural and linguistic experience of the listener (Iversen et al. 2008).

In sum, patterns of change in acoustic dimensions, such as frequency, duration, and intensity produce subjective accents on musical elements and influence how those elements are perceived to be grouped together in time, permitting a figural coding of rhythm. Grouping can also be imposed by the listener or emphasized by a performer, and is in addition impacted by the linguistic experience of the listener. In the absence of explicit acoustic cues, listeners tend to hear groups of twos, threes, and fours; preferred group size, however, is influenced by tempo. Section 6.5.2 turns to a discussion of central issues in research on beat and meter, including different types of accents, effects of tempo, and contributions of listener knowledge and experience.

6.5.2 *Perception of Beat and Meter*

Beat and meter have been studied using a variety of different empirical methods. This work has shown that metric coding of rhythms confers a number of advantages to listeners. Critically, metric coding provides information about the relative timing of elements of a rhythm and a basis for generating expectations about “when” in time future events will occur. Rhythms that can be described by a metric hierarchy (i.e., they afford a metric coding) have been shown to be more easily discriminated than rhythms that do not fit within a metric framework (Bharucha and Pryor 1986). Listeners also have more trouble discriminating pitches that occur at metrically weak locations than those that occur at metrically strong locations (Jones et al. 1982). Moreover, memory for the temporal position of a probe tone is better when probe tones are placed at strong metrical positions than at weak metrical positions in an implied metric hierarchy (Palmer and Krumhansl 1990). Finally, beat and meter facilitate judgments about time and pitch changes (Jones et al. 1982, 2002; Jones and Yee 1997; Barnes and Jones 2000; McAuley and Jones 2003) and enhance judgments of event durations (Boltz 1991, 1998) and melodic phrase completeness (Boltz 1989).

6.5.2.1 *Contribution of Different Types of Accents*

An important theoretical issue in work on beat and meter concerns the contribution of different types of accents and their timing to the communication of metrical structure to a listener (Cooper and Meyer 1960; Benjamin 1984). Two kinds of accents that have received substantial consideration in the literature are temporal accents and melodic accents. Temporal accents include pause accents and duration accents. Pause (or rhythmic) accents are produced by an empty time interval (marked by the onsets of two successive tones) that is relatively long compared to preceding inter-onset intervals (Povel and Okkerman 1981; Narmour 1996; Jones 1987; Jones and Pfordresher 1997). Duration accents are accents that occur on tones with a relatively long duration compared to the duration of preceding tones (Woodrow 1951; Handel 1989).

There are several varieties of melodic accents (Jones 1993; Hannon et al. 2004). Interval accents are created on a tone when the tone is much higher or lower in pitch than the surrounding events (Lerdahl and Jackendoff 1983; Huron and Royal 1996). With an interval accent, it is the element after the pitch jump (from high-to-low or low-to-high) that is accented (Jones 1981, 1993). Contour accents occur on tones at the point of change in a musical contour (e.g., the middle tone in three-tone melody than ascends and then descends in pitch); these points have been called contour pivot points or turnaround points (Thomassen 1982). It is not surprising that interval and contour accents frequently overlap because turnaround points often involve large pitch leaps. A third instance of melodic accent is a tonal accent (Smith and Cuddy 1989; Dawe et al. 1993). Tonal accents arise from a shift in tonal stability (e.g., from a leading tone to the tonic) within a particular musical context.

One issue that has received increased attention recently concerns the relative weight that listeners place on melodic and temporal accents in inferring metrical structure (Ellis and Jones 2009). Although there is consistent support that temporal accents are very important for metric coding (Povel and Essens 1985; Large and Jones 1999), data on the importance of melodic accents for perceiving metrical structure are mixed (Huron and Royal 1996). On the one hand, a number of studies have shown relatively little contribution of melodic accents to meter. Snyder and Krumhansl (2001) showed that when participants were asked to tap to pitch-varied and monotone versions of ragtime piano music, there was very little difference in tapping performance for the two conditions. Woodrow (1911) showed that increases in loudness and duration, but not changes in pitch influence the perceived beginning of a group of elements. Metrical stability ratings for events in melodies interrupted at various points show larger effects for temporal accents than for pitch accents (Bigand 1997). Finally, in expressive musical performance, pitch accents tend to be less consistent than temporal accents and highly context dependent (Drake and Palmer 1993).

On the other hand, there is also empirical evidence that both melodic accents and temporal accents contribute significantly to the perception of metrical structure; see Ellis and Jones (2009) for a comprehensive review. Much of this work highlights the importance of the temporal alignment of the two types of accents. This view is most comprehensively expressed by entrainment-based approaches, such as Dynamic Attending Theory (DAT; Jones 1976; Jones and Boltz 1989; Large and Jones 1999). In DAT, the periodic timing of temporal accents and melodic accents is assumed to drive entrainment and contribute to the emergence of what Jones and colleagues refer to as joint accent structure. A key prediction of this theory is that coincident (concordant) melodic and temporal accents (corresponding to a simple joint accent structure) should lead more efficient entrainment and a stronger perception of metrical structure than conflicting (discordant) accent timing (corresponding to a complex joint accent structure). Empirical support for a joint accent structure (JAS) hypothesis has been found in a number of studies using a variety of tasks. In general, melodies with a concordant JAS have been shown to produce perceptual and performance advantages over melodies with a discordant JAS (Deutsch 1980; Boltz and Jones 1986; Dowling et al. 1987; Monahan et al. 1987; Boltz 1989, 1991; Drake et al. 1991; Jones et al. 1993; Jones and Pfordresher 1997; Pfordresher 2003).

Overall, conflicting results concerning the importance of melodic accents to meter perception raise the obvious question of why findings across studies have been so inconsistent. One possible reason is a failure of most studies to control for accent salience (Huron and Royal 1996; Snyder and Krumhansl 2001; Temperley and Bartlette 2002; Toivianen and Snyder 2003). Thus, in cases where melodic accents are likely to be less salient than temporal accents, it is perhaps not surprising that the data favor temporal accents with melodic accents contributing weakly or not at all to perceived meter. This suggests that the relative contribution of melodic and temporal accents can be carefully assessed only when melodic and temporal accents are equated for their salience. In one of the only studies to explicitly equate melodic and temporal salience, Ellis and Jones (2009) provide conclusive evidence that melodic accents do contribute to meter perception and that metrical clarity ratings are greater

for melodies with a concordant JAS than for melodies with a discordant JAS. Moreover, metrical clarity ratings are found to increase and reaction times decrease as the number of temporally aligned (coincident) accents increases.

6.5.2.2 Role of Tempo in Perception of Metrical Structure

As with many aspects of rhythm perception, tempo plays an important supporting role. This is particularly true for the perception of beat and meter, where the tempo of a rhythm can influence how listeners organize the elements of the rhythm both in terms of figural coding and in terms of metric coding (London 2004). The focus of this section is on the role of tempo in metric coding. One common empirical method for studying effects of tempo on perceived beat and meter requires listeners to synchronize taps with various rhythms presented at different tempi. A key dependent measure in studies using this method is the level of the implied metric hierarchy that listeners decide to tap.

In an influential set of studies on this topic, Handel and colleagues (Oshinsky and Handel 1978; Handel and Oshinsky 1981; Handel and Lawson 1983; Handel 1984) had listeners tap in synchrony with what they perceived to be the most natural placement of accents for a variety of constructed polyrhythms. Polyrhythms pit one isochronous rhythm against another, and so naturally create several possible rhythm interpretations of the emerging pattern. For example, for a 3×4 polyrhythm, Oshinsky and Handel (1978) observed three different tapping responses depending on tempo: (1) Listeners subdivided the pattern into three equal time intervals, tapping in synchrony with the three-element sequence; (2) they subdivided the pattern into four equal time intervals, tapping in synchrony with the four-element sequence; or (3) they tapped once every 12 elements when the two sequences coincided. Overall, listeners were more likely to subdivide into three equal time intervals at fast tempi than at slow tempi. Handel and colleagues have reported similar effects for other polyrhythms.

Related research has shown that for isochronous rhythms, individuals tend to tap beats every two, three, or four elements in a tempo-dependent fashion (Duke 1989); this finding is reminiscent of work by Bolton (1894) on grouping, but differs in that listeners are not asked to explicitly group elements, but rather to tap what they perceived to be the beat of the rhythm. Still other work has shown that for both simple and complex meters characteristic of Western music, there is a tendency to tap out higher time levels in an implied metric hierarchy at fast tempos than at slow tempos (Duke 1989; Parncutt 1994; London 2002, 2004). In many cases at very fast tempos, there is a tendency for individuals to tap once with each repetition of the rhythm (the highest level in the hierarchy; McAuley and Semple 1999).

In sum, for variety of rhythms, the most salient time level (i.e., the perceived beat) for listeners tends to be at a higher level in the metric hierarchy at fast tempi than at slow tempi. One interpretation of the reported effects of tempo on perceived beat and meter is that there is an interaction between preferred tempo and metrical structure. As the tempo of a rhythm changes, the relative time level that is nearest

to an absolute preferred time interval also changes; in turn, this affects the relative salience of each level of the hierarchy. Having considered the role of different type of accents and the contributions of tempo to beat and meter, the discussion turns to the role of knowledge and experience in the next section.

6.5.2.3 Role of Knowledge and Experience in Perception of Metrical Structure

Elements of beat and meter perception are present relatively early in infancy (Hannon and Johnson 2005). Nonetheless, perception of beat and meter is also subject to learning and effects of enculturation (Hannon and Trehub 2005); see Trainor and Corrigan, Chap. 4, for a developmental perspective on these issues. In adulthood, questions about the role of knowledge in the perception of metrical structure have centered on the type of knowledge that musically trained and untrained individuals bring to bear when listening to music. That is, to what extent does meter have a psychological basis and do both trained and untrained listeners apply their knowledge of meter in everyday musical listening situations?

In a seminal study on this topic, Palmer and Krumhansl (1990) provided several sources of evidence that support mental representations for metrical structure and its use in perception. First, analysis of the frequency distribution of notes in a corpus of musical excerpts revealed statistical regularities in the temporal distribution of musical events that permitted accurate identification of the meter for at least the selected subset of meters examined in the study. Second and more important, in the absence of any acoustical cues to meter, listener judgments about the goodness-of-fit of the timing of probe tones placed at different metrical positions paralleled the pattern of statistical regularities observed in the corpus analyses. Third, effects of musical training were also evident with musicians showing finer-grained representations of metrical structure than did nonmusicians. Finally, listeners' memory for the temporal placement of a probe tone was influenced by its position in the implied metrical hierarchy with better memory for probes at strong metrical positions than at weak metrical positions. Support for the mental representation of musical meter is also evident in expressive musical performance. For example, pianists use their knowledge of the noted meter to accent events at strong metrical positions by making them longer, louder, or more legato (Shaffer 1981; Sloboda 1983; Drake and Palmer 1993).

Having now identified key theoretical issues and empirical findings in research on musical rhythm, including work on grouping, beat, and meter, a final section on rhythm provides a more detailed consideration of several influential modeling approaches.

6.5.3 Models of Rhythm

An important distinction made in the development of various models of rhythm is between the processing of rhythms that permit a metric coding and those that only

afford a figural coding (Essens and Povel 1985; Povel and Essens 1985; Hébert and Cuddy 2002). Povel and Essens (1985) refer to the former metrical temporal patterns and to the later as nonmetrical temporal patterns. Much of the work on modeling rhythm has focused on the induction of a beat and the perception of metrical structure (i.e., metric coding). These models range from algorithmic rule-based approaches (e.g., Longuet-Higgins and Lee 1982; Povel and Essens 1985; Desain and Honing 1999) to dynamical systems accounts (e.g., Large and Kolen 1994; Large and Jones 1999; Eck 2002).

In general, rule-based models of rhythm share similarities with interval theories of tempo. Specifically, as with tempo, rule-based models of rhythm tend to derive from an information processing perspective with the common assumption that people perceive, remember, and reproduce rhythms by structuring their mental representation according to an internal clock. As with tempo models, this internal clock is assumed to involve a pacemaker-accumulator mechanism that ticks out regular intervals that are aligned with particular stimulus onsets that correspond to induced beats.

One particularly influential rule-based approach is that of Povel and Essens (1985). These authors proposed what amounts to a three-stage clock model. First, subjective accents were assigned to the rhythm according to a set of empirically derived preference rules (Povel and Okkerman 1981). The emphasis of these preference rules was on temporal accents. Accents were assumed to occur on (a) temporally isolated tones, (b) the second in a group of two tones, and (c) the first and last tone in a run or three or more elements (Povel and Okkerman 1981; Povel and Essens 1985). Second, all possible clock intervals were generated in an algorithmic fashion, allowing some restrictions on what would be considered viable responses. Finally, in a “matching” stage, the amount of counter (negative) evidence was calculated for each potential clock and the clock with the least counter-evidence was determined to be the most likely induced beat. Negative evidence consisted of clock pulses (beats) falling on unaccented elements or silent elements of a sequence, with silent elements contributing more negative evidence than unaccented elements. In support of their model, Povel and Essens found that rhythms with less negative evidence were reproduced more accurately and judged to be simpler than rhythms with more negative evidence (Povel and Essens 1985; Essens and Povel 1985).

McAuley and Semple (1999) generalized the Povel and Essens (1985) clock model approach by in addition considering models that determined the best clock (beat) according to a positive evidence heuristic, as well as a hybrid model that combined in a weighted fashion both positive and negative (counter-evidence) evidence heuristics. The use of positive evidence as a heuristic has been similarly examined by Parncutt (1994), who also incorporated into his model the concept of pulse salience in an attempt to address effects of tempo. Because the three models differ in the nature of the evidence used to determine the strength of the best-fitting clock (beat), they each have different biases. Clock models involving negative evidence heuristics tend to favor slow clocks because slow clocks afford less opportunity to accumulate negative evidence. In contrast, clock models involving positive evidence heuristics tend to favor fast clocks because fast clocks afford

more opportunity to accumulate positive evidence. Hybrid models that combine positive and negative evidence in a weighted fashion tend to offer a balance between these two extremes and favor clocks that are neither too fast nor too slow.

In a comparison of these three models (positive evidence, negative evidence, hybrid), McAuley and Semple (1999) found that models based solely on a negative-evidence heuristic best predicted the perceived beat of nonmusicians, whereas models incorporating a positive-evidence heuristic best predicted the perceived beat of musicians. In addition, tempo tended to affect the type of evidence that best predicted the perceived beat. A negative-evidence heuristic worked best for fast tempi because it favored, like listeners, beats at higher levels of the metric hierarchy. In contrast, a positive-evidence heuristic worked best for slow tempi because it favored, like listeners, beats at lower levels of the metric hierarchy. The tempo shift favoring negative-evidence heuristics at fast tempi and positive-evidence heuristics at a slow tempi reflects listeners tendency to prefer to tap beats at a level within a metric hierarchy that it is at an absolute tempo that is neither too fast nor too slow, but rather falls at an intermediate rate. This intermediate rate is in the range of commonly reported preferred tempi.

Three general weaknesses of the different varieties of the internal clock approach to rhythm are that these models: (1) describe, but fail to explain why various factors such as tempo and musical experience alter the perception of a beat; (2) do not operate in real time, but rather consider all possible clocks in an algorithmic fashion before settling on a solution; and (3) rely primarily on temporal accents to infer beat and meter. In addition to the important role that different types of accents play, another somewhat overlooked factor is the role of repetition. Both the repetition of melodic fragments and rhythmic patterns contribute to the perception of metrical structure (Steedman 1977; Temperley 2001). See research by Steedman (1977) and Temperley and Bartlette (2002) for examples of models incorporating a principle of repetition (or parallelism); these models have been used to recover the scored meter of a piece of music with modest success.

A variety of models have been proposed that do operate in real time, many arising from an entrainment perspective. Most entail hypotheses about self-sustaining oscillations that respond to both the timing and intensity of event onsets and facilitate the extraction of a musical pulse, which is then used for real-time beat tracking (Large and Kolen 1994; McAuley 1995; Toiviainen 1998; Eck 2002; Large and Palmer 2002). The perception of meter from an entrainment perspective requires a multiple-oscillator model in which each oscillator corresponds to each metrical level. One proposal along these lines is for the relative salience of each metrical level to be modeled by the resonance region of an oscillator that is centered on preferred tempo (van Noorden and Moelants 1999).

Overall, entrainment-based approaches to beat and meter offer advantages over internal clock approaches in that they operate in real time and have the potential for providing a more comprehensive explanation of how factors such as tempo influence beat perception and meter (Large and Kolen 1994; McAuley 1995; Large and Jones 1999). As with interval models, entrainment approaches are similarly limited in their reliance primarily on temporal accents to infer beat and meter.

6.6 Summary

Tempo and rhythm are two fundamental and related aspects of musical communication. Tempo refers to pace of a piece of music (i.e., how fast or slow it is) and is typically measured in beats per minute or by the beat-onset interval. The concept of rhythm has been less straightforward to define and can be viewed either as a serial pattern of durations or the perceived temporal organization of that pattern. In the second sense of the term, rhythm as a perception has several elements including grouping, beat, and meter.

Theoretical approaches to tempo and rhythm can, for the most part, be classified as either interval theories or entrainment (beat-based) theories. Key differences between the two approaches concern (1) the nature of the clock's responses to stimulus onsets (arbitrary reset via a switch in interval theories vs. more gradual phase and period correction processes in entrainment theories), (2) representation of duration (duration code stored in memory in interval theories vs. oscillator period in entrainment theories), (3) successive time estimates (independent vs. dependent), and (4) the nature of duration judgments (explicit comparison of two stored duration codes in interval theories vs. temporal contrast metric in entrainment theories).

Perceptual and motor preferences for particular tempi emerge at a young age. In adulthood, representative values for preferred perceptual tempo and spontaneous motor tempo are around 600 ms, although correlations between the two are modest. Across the life span, there is converging support for the view that preferred tempo slows with increased age. There is also support for the view that listeners develop absolute memory for tempi of music that is experienced at the same tempo (i.e., the music has a tempo standard). There is little support, however, for differences in preferred tempo associated with gender, handedness, or for clear associations with specific physiological measures. Experience does appear to play a role as musicians tend to have slower preferred tempi than nonmusicians, especially in childhood.

Studies of tempo discrimination and tempo production reveal violations in Weber's law consistent with the concept of a preferred tempo. With respect to perception, JNDs for tempo tend to be a minimum in a range of tempi (optimal tempo region) that is centered on 600 ms, with thresholds for single interval sequences around 6% and those for multiple-interval isochronous sequences around 2% in ideal listening conditions. Weber fractions for variability in tempo productions tend to mirror threshold results observed for perception. There is also converging evidence both musicians and nonmusicians develop fairly precise memories for the absolute tempo of music, as long as the music has a tempo standard

Research on rhythm distinguishes the perception of grouping (figural coding) from the perception of beat and meter (metric coding). Some of the earliest work on rhythm showed that for isochronous sequences of identical sounds, listeners tend to perceive the elements of the sequences to be grouped in twos, threes, or fours, with the first element of each group judged to be more accented than the others. Acoustic factors affecting perceived grouping (i.e., figural coding of a rhythm)

include intensity, duration, and frequency patterning. The topic of beat and meter has been studied using a variety of different methods, with the general finding that musical events that occur at strong metrical positions are better perceived, remembered, and reproduced than musical events that occur at weak metrical positions. Both stimulus-based accents of different types and musical knowledge contribute to perceived beat and meter (i.e., metric coding of a rhythm). There is consistent support that the timing of temporal accents is important for the perception of metrical structure, but the role of melodic accents in metric coding is less clear cut. However, many studies revealing weak or negligible effects of melodic accents on perceived beat and/or meter have failed to control for accent salience.

Overall, interval and entrainment models of tempo and rhythm have met with mixed success. Interval models of tempo, such as the multiple-look and generalized multiple-look models have been quite successful in describing discrimination thresholds for tasks requiring the discrimination of isolated time intervals or isochronous sequences. However, these approaches have been less successful when applied to more complex rhythms. Moreover, because the pacemaker-accumulator conception of an internal clock passively records time as the number of ticks, interval models, in general, are agnostic about the concept of a preferred tempo. Entrainment models offer an alternative approach, which addresses both findings on preferred tempo and tempo discrimination.

One outstanding challenge of many models of rhythm is to address the finding that the perceived grouping, beat and meter (and hence figural and metric coding) vary with tempo, as well as vary from listener to listener. Another issue that has been difficult to address in models is the nature of interactions between pitch and time cues in perception of grouping, beat, and meter. Most models of rhythm have focused on the relative timing of temporal accents; relatively few, in contrast, address the role of melodic accents. Dynamic attending theory developed within an entrainment framework is one approach that shows promise for capturing the rich interactions between melodic and temporal accents in the perception of rhythm.

Finally, a generally understudied area of research is individual differences. With respect to rhythm, there are numerous anecdotal reports that suggest that there are large individual differences in the ability to perceive a beat; that is, some people appear to have much more difficulty perceiving a beat than others. The nature of these individual differences in adulthood is only beginning to be addressed (Iversen and Patel 2008; Grahn and McAuley 2009). The importance of this line of investigation is highlighted by recent evidence supporting a link between rhythmic ability and language processing (Alcock et al. 2000; Thompson and Goswami 2008) and evidence of rhythm perception deficits in neurological disorders, such as Parkinson's disease (Grahn and Brett 2009).

Acknowledgements Many thanks to Mari Riess Jones, Laura Dille, Molly Henry, and Nathan Miller for detailed comments on an earlier version of this chapter. Thanks also to Laura Dille for help with figures. The author received support for this work from the National Science Foundation (BCS 0818271).

References

- Abel SM (1972) Discrimination of temporal gaps. *J Acoust Soc Am* 52:519–524.
- Alcock KJ, Passingham RE, Watkins KE, Vargha-Khadem F (2000) Pitch and timing abilities in inherited speech and language impairment. *Brain Lang* 75:34–46.
- Allan LG (1979) The perception of time. *Percept Psychophys* 26:340–354.
- Andrews MW, Dowling WJ, Bartlett JC, Halpern AR (1998) Identification of speeded and slowed familiar melodies by younger, middle-aged, and older musicians and nonmusicians. *Psychol Aging* 13:462–471.
- Bamberger J (1980) Cognitive structuring in the apprehension and description of simple rhythms. *Arch Psychol* 48:171–199.
- Barnes R, Jones MR (2000) Expectancy, attention, and time. *Cogn Psychol* 41:254–311.
- Benjamin WE (1984) A theory of musical meter. *Music Percept* 1:355–413.
- Bergeson T, Trehub S (2002) Absolute pitch and tempo in mothers' songs to infants. *Psychol Sci* 13:72–75.
- Bharucha J, Pryor JH (1986) Disrupting the isochrony underlying rhythm: an asymmetry in discrimination. *Percept Psychophys* 40:137–141.
- Bigand E (1997) Perceiving musical stability: the effect of tonal structure, rhythm, and musical expertise. *J Exp Psychol: Hum Percept Perform* 23:808–822.
- Bolton TL (1894) Rhythm. *Am J Psychol* 6:145–238.
- Boltz M (1989) Rhythm and “good endings”: effects of temporal structure on tonality judgments. *Percept Psychophys* 46:9–17.
- Boltz M (1991) Some structural determinants of melody recall. *Mem Cogn* 19:239–251.
- Boltz M (1994) Changes in internal tempo and effects on the learning and remembering of event durations. *J Exp Psychol: Learning Mem Cogn* 20:1154–1171.
- Boltz M (1998) The processing of temporal and nontemporal information in the remembering of event durations and musical structure. *J Exp Psychol: Hum Percept Perform* 24:1087–1104.
- Boltz M, Jones MR (1986) Does rule recursion make melodies easier to reproduce? If not, what does? *Cogn Psychol* 18:389–431.
- Church RM (2003) A concise introduction to scalar timing theory. In Meck WH (ed), *Functional and Neural Mechanisms of Interval Timing*. Boca Raton, FL: CRC Press.
- Clarke E (1999) Rhythm and timing in music. In Deutsch D (ed), *The Psychology of Music*, 2nd ed. New York: Academic Press, pp. 473–500.
- Collyer CE, Broadbent HA, Church, RM (1994) Preferred rates of repetitive tapping and categorical time production. *Percept Psychophys* 55:443–453.
- Cooper GW, Meyer LB (1960) *The Rhythmic Structure of Music*. Chicago, IL: University of Chicago Press.
- Creelman CD (1962) Human discrimination of auditory duration. *J Acoust Soc Am* 34:582–593.
- Crocker RL (2000) *An Introduction to Gregorian Chant*. New Haven, CT: Yale University Press.
- Dalla Bella S, Peretz I, Rousseau L, Gosselin N (2001) A developmental study of the affective value of tempo and mode in music. *Cognition* 80:1–10.
- Dawe LA, Platt JR, Racine RJ (1993) Harmonic accents in inference of metrical structure and perception of rhythm patterns. *Percept Psychophys* 54:794–807.
- Desain P, Honing H (1999) Computational models of beat induction: the rule-based approach. *J New Music Res* 28:29–42.
- Deutsch D (1980) The processing of structured and unstructured tonal sequences. *Percept Psychophys* 28:381–389.
- Dowling WJ, Lung KM, Herrbold S (1987) Aiming attention in pitch and time in the perception of interleaved melodies. *Percept Psychophys* 41:642–656.
- Drake C (1993) Perceptual and performed accents in musical sequences. *Bull Psychonomic Soc* 31:107–110.
- Drake C, Botte MC (1993) Tempo sensitivity in auditory sequences: evidence for a multiple-look model. *Percept Psychophys* 54:277–286.

- Drake C, Palmer C (1993) Accent structure in music performance. *Music Percept* 10:343–378.
- Drake C, Dowling WJ, Palmer C (1991) Accent structures in the reproduction of simple tunes by children and adult pianists. *Music Percept* 8:315–334.
- Drake C, Jones MR, Baruch C (2000) The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. *Cognition* 77:251–288.
- Duchek JM, Balota DA, Ferraro FR (1994) Component analysis of a rhythmic finger tapping task in individuals with senile dementia of the Alzheimer type and in individuals with Parkinson's disease. *Neuropsychology* 8:218–226.
- Duke RA (1989) Musicians' perception of beat in monotonic stimuli. *J Res Music Ed* 37:61–71.
- Eck D (2002) Finding downbeats with a relaxation oscillator. *Psychol Res* 66:18–25.
- Ellis RJ, Jones MR (2009) The role of accent salience and joint accent structure in meter perception. *J Exp Psychol: Hum Percept Perform* 35:264–280.
- Essens PJ, Povel DJ (1985) Metrical and nonmetrical representations of temporal patterns. *Percept Psychophys* 37:1–7.
- Farnsworth P, Block H, Waterman W (1934) Absolute tempo. *J Gen Psychol* 10:230–233.
- Fraisse P (1956) *Les structures rythmiques*. Louvain, Belgium: Publications Universitaires de Louvain.
- Fraisse P (1963) *The Psychology of Time*. New York: Harper and Row.
- Fraisse P (1982) Rhythm and tempo. In Deutsch D (ed), *The Psychology of Music*. Orlando, FL: Academic, pp. 149–180.
- Friberg A, Sundberg, J (1995) Time discrimination in a monotonic, isochronous sequence. *J Acoust Soc Am* 98:2524–2531.
- Friberg A, Sundström (2002) Swing ratios and ensemble timing in jazz performance: evidence for a common rhythmic pattern. *Music Percept* 19:333–349.
- Frischeisen-Köhler, I (1933) The personal tempo and its inheritance. *Character Pers* 1:301–313.
- Gagnon L, Peretz I (2003) Mode and tempo relative contributions to “happy-sad” judgements in equitone melodies *Cogn Emotion* 17:25–40.
- Garner WR (1974) *The Processing of Information and Structure*. Oxford: Lawrence Erlbaum.
- Getty DJ (1975) Discrimination of short temporal intervals: a comparison of two models. *Percept Psychophys* 18:1–8.
- Geuze RH, Kalverboer, AF (1994) Tapping a rhythm: a problem of timing for children who are clumsy and dyslexic? *Adapt Phys Act Q* 11:203–213.
- Gibbon J (1977) Scalar expectancy theory and Weber's law in animal timing. *Psychol Rev* 84:279–325.
- Gibbon J, Church RM, Meck WH (1984) Scalar timing in memory. *Ann NY Acad Sci* 423:52–77.
- Grahn JA, Brett M (2009) Impairment of beat-based rhythm discrimination in Parkinson's disease. *Cortex* 45:54–61.
- Grahn JA, McAuley JD (2009) Neural bases of individual differences in beat perception. *NeuroImage* 47:1894–1903.
- Greene LS, Williams, HG (1993) Age-related differences in timing control of repetitive movement: application of the Wing-Kristofferson model. *Res Q Exerc Sport* 64:32–38.
- Grondin S (2001a) Discriminating time intervals presented in sequences marked by visual signals. *Percept Psychophys* 63:1214–1228.
- Grondin S (2001b) From physical time to the first and second moments of psychological time. *Psychol Bull* 47:22–44.
- Grondin S, McAuley JD (2009) Duration discrimination in crossmodal sequences. *Perception* 38:1542–1559.
- Halpern AR (1988) Perceived and imagined tempos of familiar songs. *Music Percept* 6:193–202.
- Handel S (1984) Using polyrhythms to study rhythm. *Music Percept* 1:465–484.
- Handel S (1989) *Listening: An Introduction to the Perception of Auditory Events*. Cambridge, MA: MIT Press.
- Handel S (1998) The interplay between metric and figural rhythmic organization. *J Exp Psychol: Hum Percept Perform* 24:1546–1561.

- Handel S, Lawson GR (1983) The contextual nature of rhythmic interpretation. *Percept Psychophys* 34:103–120.
- Handel S, Oshinsky, JS (1981) The meter of syncopated auditory polyrhythms. *Percept Psychophys* 30:1–9.
- Hannon EE, Johnson SP (2005) Infants use meter to categorize rhythms and melodies: implications for musical structure learning. *Cogn Psychol* 50:354–377.
- Hannon EE, Trehub SE (2005) Metrical categories in infancy and adulthood. *Psychol Sci* 16:48–55.
- Hannon EE, Snyder JS, Eerola T, Krumhansl CL (2004) The role of melodic and temporal cues in perceiving musical meter. *J Exp Psychol: Hum Percept Perform* 30:956–974.
- Harrel TW (1937) Factors influencing preference and memory for auditory rhythm. *J Gen Psychol* 17:63–104.
- Hébert S, Cuddy LL (2002) Detection of metric structure in auditory figural patterns. *Percept Psychophys* 64:909–918.
- Hirsh IJ, Monahan CB, Grant KW, Singh PG (1990) Studies in auditory timing: 1. Simple patterns. *Percept Psychophys* 47:215–226.
- Huron D, Royal M (1996) What is melodic accent? Converging evidence from musical practice. *Music Percept* 13:489–516.
- Iversen JR, Patel AD (2008) The beat alignment test (BAT): surveying beat processing abilities in the general population. In Adachi et al. (eds), *Proceedings of the 10th International Conference on Music Perception and Cognition*, August 25–29. Adelaide: Causal Productions.
- Iversen JR, Patel AD, Ohgushi K (2008) Perception of rhythmic grouping depends on auditory experience. *J Acoust Soc Am* 124:2263–2271.
- Ivry RB, Hazeltine RE (1995) Perception and production of temporal intervals across a range of durations: evidence for a common timing mechanism. *J Exp Psychol: Hum Percept Perform* 21:3–18.
- Ivry RB, Keele SW (1989) Timing functions of the cerebellum. *J Cogn Neurosci* 6:136–152.
- James W (1890) *The Principles of Psychology*, Vol. I. New York: Holt.
- Jones MR (1976) Time, our lost dimension: toward a new theory of perception, attention, and memory. *Psychol Rev* 83:323–355.
- Jones MR (1981) A tutorial on some issues and methods in serial pattern research. *Percept Psychophys* 30:492–504.
- Jones MR (1987) Dynamic pattern structure in music: recent theory and research. *Percept Psychophys* 41:621–634.
- Jones MR (1993) Dynamics of musical patterns: how do melody and rhythm fit together? In Tighe TJ, Dowling WJ (eds), *Psychology and Music: The Understanding of Melody and Rhythm*. Hillsdale, NJ: Lawrence Erlbaum, pp. 67–92.
- Jones MR, Boltz M (1989) Dynamic attending and responses to time. *Psychol Rev* 96:459–491.
- Jones MR, McAuley JD (2005) Time judgments in global temporal contexts. *Percept Psychophys* 67:398–417.
- Jones MR, Pfordresher PQ (1997) Tracking melodic events using joint accent structure. *Can J Exp Psychol* 51:271–291.
- Jones MR, Ralston JT (1991) Some influences of accent structure on melody recognition. *Mem Cogn* 19:8–20.
- Jones MR, Yee W (1997) Sensitivity to time change: the role of context and skill. *J Exp Psychol: Hum Percept Perform* 23:693–709.
- Jones MR, Kidd GR, Wetzel R (1981) Evidence for rhythmic attention. *J Exp Psychol: Hum Percept Perform* 7:1059–1073.
- Jones MR, Boltz M, Kidd GR (1982) Controlled attending as a function of melodic and temporal context. *Percept Psychophys* 32:211–218.
- Jones MR, Boltz M, Klein JM (1993) Expected endings and judged duration. *Mem Cogn* 21:646–665.
- Jones MR, Moynihan H, Mackenzie N, Puente J (2002) Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychol Sci* 13:313–319.
- Keele SW, Nicoletti R, Ivry RI, Pokorny RA (1989) Mechanisms of perceptual timing: beat-based or interval-based judgments? *Psychol Res* 50:221–256.

- Klein JM, Jones MR (1996) Effects of attentional set and rhythmic complexity on attending. *Percept Psychophys* 58:34–46.
- Krampe R, Mayr U, Kliegl R (2005) Timing, sequencing, and executive control in repetitive movement production. *J Exp Psychol: Hum Percept Perform* 31:379–397.
- Large EW, Jones MR (1999) The dynamics of attending: how people track time-varying events. *Psychol Rev* 107:119–159.
- Large EW, Kolen JF (1994) Resonance and the perception of musical meter. *Connection Sci* 6:177–208.
- Large EW, Palmer C (2002) Perceiving temporal regularity in music. *Cogn Sci* 26:1–37.
- Large EW, Fink P, Kelso JAS (2002) Tracking simple and complex sequences. *Psychol Res* 66:3–17.
- Lerdahl F, Jackendoff R (1983) *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.
- Levitin DJ, Cook PR (1996) Memory for musical tempo: additional evidence that auditory memory is absolute. *Percept Psychophys* 58:927–935.
- London J (2002) Cognitive constraints on metric systems: some observations and hypotheses. *Music Percept* 19:529–550.
- London J (2004) *Hearing in Time: Psychological Aspects of Musical Meter*. New York: Oxford University Press.
- Longuet-Higgins C, Lee C (1982) The perception of musical rhythms. *Perception* 11:115–128.
- Mattel MS, Meck WH (2000) Neuropsychological mechanisms of interval timing behavior. *BioEssays* 22:94–103.
- McAuley JD (1995) Perception of time as phase: toward an adaptive-oscillator model of rhythmic pattern processing. Unpublished doctoral dissertation, Indiana University, Bloomington.
- McAuley JD, Jones MR (2003) Modeling effects of rhythmic context on perceived duration: a comparison of interval and entrainment approaches to short-interval timing. *J Exp Psychol: Hum Percept Perform* 29:1102–1125.
- McAuley JD, Kidd GR (1998) Effect of deviations from temporal expectations on tempo discrimination of isochronous tone sequences. *J Exp Psychol: Hum Percept Perform* 24:1786–1800.
- McAuley JD, Semple P (1999) The effect of tempo and musical experience on perceived beat. *Aust J Psychol* 51:176–187.
- McAuley JD, Jones MR, Holub S, Johnston HM, Miller NS (2006) The time of our lives: lifespan development of timing and event tracking. *J Exp Psychol: Gen* 135:348–367.
- Meck WH (1996) Neuropharmacology of timing and time perception. *Cogn Brain Res* 3:227–242.
- Meck WH (2003) *Functional and Neural Mechanisms of Interval Timing*. Boca Raton, FL: CRC Press.
- Meck WH (2005) Neuropsychology of timing and time perception. *Brain Cogn* 58:1–8.
- Michon JA (1964) Studies on subjective duration: I. Differential sensitivity in the perception of repeated temporal intervals. *Acta Psychol* 22:441–450.
- Miller NS, McAuley JD (2005) Tempo sensitivity in isochronous tone sequences: the multiple-look model revisited. *Percept Psychophys* 67:1150–1160.
- Mishima J (1956) On the factors of mental tempo. *Japanese Psychol Res* 4:27–38.
- Monahan CB, Kendall RA, Carterette EC (1987) The effect of melodic and temporal contour on recognition memory for pitch change. *Percept Psychophys* 41:576–600.
- Moore-Ede MC, Sulzman FM, Fuller CA (1982) *The Clocks that Time Us: Physiology of the Circadian Timing System*. Cambridge, MA: Harvard University Press.
- Narmour E (1996) Analyzing form and measuring perceptual content in Mozart's Sonata K. 282: a new theory of parametric analogues. *Music Percept* 13:728–741.
- Ogden RT, Collier GL (1999) On detecting and modeling deterministic drift in long run sequences of tapping data. *Comm Stat Theory Methods* 28:977–987.
- Oshinsky JS, Handel S (1978) Syncopated auditory polyrhythms: discontinuous reversals in meter interpretation. *J Acoust Soc Am* 63:936–939.
- Palmer C, Krumhansl CL (1990) Mental representations of musical meter. *J Exp Psychol: Hum Percept Perform* 16:728–741.

- Parncutt R (1994) A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Percept* 11:409–464.
- Pashler H (2001) Perception and production of brief durations: beat-based versus interval-based timing. *J Exp Psychol: Hum Percept Perform* 27:485–493.
- Patel AD, Iversen JR, Chen Y, Repp BH (2005) The influence of metricality and modality on synchronization with a beat. *Exp Brain Res* 163:226–238.
- Pfordresher, PQ (2003) The role of melodic and rhythmic accents in musical structure. *Music Percept* 20:431–464.
- Pöppel E (1997) A hierarchical model of temporal perception. *Trends Cogn Sci* 1:56–61.
- Povel DJ, Essens PJ (1985) Perception of temporal patterns. *Music Percept* 2:411–440.
- Povel DJ, Okkerman H (1981) Accents in equitone sequences. *Percept Psychophys* 30:565–572.
- Repp BH (2003) Rate limits in sensorimotor synchronization with auditory and visual sequences: the synchronization threshold and the benefits and costs of interval subdivision. *J Mot Behav* 35:355–370.
- Repp BH (2005) Rate limits of on-beat and off-beat tapping with simple auditory rhythms: 2. The roles of different kinds of accent. *Music Percept* 23:165–187.
- Repp BH, Windsor WL, Desain P (2002) Effects of tempo on the timing of simple musical rhythms. *Music Percept* 19:565–593.
- Rimoldi HJA (1951) Personal tempo. *J Abnorm Soc Psychol* 46:280–303.
- Ross J, Houtsma AJM (1994) Discrimination of auditory temporal patterns. *Percept Psychophys* 56:19–26.
- Rousseau L, Rousseau R (1996) Stop-reaction time and the internal clock. *Percept Psychophys* 58:434–448.
- Schulze HH (1978) The detectability of local and global displacements in regular rhythmic patterns. *Psychol Res* 40:173–181.
- Schulze HH (1989) The perception of temporal deviations in isochronic patterns. *Percept Psychophys* 45:291–296.
- Shaffer L (1981) Performances of Chopin, Bach and Bartok: studies in motor programming. *Cogn Psychol* 13:327–376.
- Sloboda JA (1983) The communication of musical metre in piano performance. *Q J Exp Psychol A* 35:377–396.
- Small AM, Campbell RA (1962) Temporal differential sensitivity for auditory stimuli. *Am J Psychol* 75:401–410.
- Smith KC, Cuddy LL (1989) Effects of metric and harmonic rhythm on the detection of pitch alterations in melodic sequences. *J Exp Psychol: Hum Percept Perform* 15:457–471.
- Smith KC, Cuddy LL, Upton R (1994) Figural and metric understanding of rhythm. *Psychol Music* 22:117–135.
- Smoll FL, Schutz RW (1978) Relationships among measures of preferred tempo and motor rhythm. *Percept Mot Skills* 46:883–94.
- Snyder JS, Krumhansl CL (2001) Tapping to ragtime: cues to pulse finding. *Music Percept* 18:455–489.
- Snyder JS, Hannon EE, Large EW, Christiansen MH (2006) Synchronization and continuation tapping to complex meters. *Music Percept* 24:135–146.
- Steedman MJ (1977) The perception of musical rhythm and metre. *Perception* 6:555–569.
- Stern W (1900) Das psychisch Tempo. In *Über psychologie der individuellen differenzen*. Leipzig: Barth.
- Stevens LT (1886) On the time sense. *Mind* 11:393–404.
- Szelag E, von Steinbüchel N, Reiser M, Gilles de Langen E, Pöppel E (1996) Temporal constraints in processing of nonverbal rhythmic patterns. *Acta Neurobiol Exp* 56:215–225.
- Temperley D (2001) *The Cognition of Basic Musical Structures*. Cambridge, MA: MIT Press.
- Temperley D, Bartlette C (2002) Parallelism as a factor in metrical analysis. *Music Percept* 20:117–149.
- ten Hoopen G, Akerboom S (1983) The subjective tempo difference between interaural and monaural sequences as a function of sequence length. *Percept Psychophys* 34:465–469.

- ten Hoopen G, Boelaarts L, Gruisen A, Apon I, Donders K, Mul N, Akerboom S (1994) The detection of anisochrony in monaural and interaural sound sequences. *Percept Psychophys* 56:110–120.
- Thomassen JM (1982) Melodic accent: experiments and a tentative model *J Acoust. Soc Am* 71:1596–1605.
- Thompson JM, Goswami U (2008) Rhythmic processing in children with developmental dyslexia: auditory and motor rhythms link to reading and spelling. *J Physiol Paris* 102:120–129.
- Toiviainen P (1998) An interactive MIDI accompanist. *Comput Music J* 22:63–75.
- Toiviainen P, Snyder JS (2003) Tapping to Bach: resonance-based modeling of pulse. *Music Percept* 21:43–80.
- van Noorden L, Moelants D (1999) Resonance in the perception of musical pulse. *J Music Res* 28:43–66.
- Vanneste S, Pouthas V, Wearden J (2001) Temporal control of rhythmic performance: a comparison between young and old adults. *Exp Aging Res* 27:83–102.
- Vierordt K (1868) *Der Zeitsinn nach Versuchen*. Tübingen: H. Laupp.
- Vos PG, van Assen M, Franek M (1997) Perceived tempo change is dependent on base tempo and direction of change: evidence for a generalized version of Schulze's (1978) internal beat model. *Psychol Res* 59:240–247.
- Wallin J (1911) Experimental studies of rhythm and time. *Psychol Rev* 18:100–133.
- Williams HG, Woollacott MH, Ivry RB (1992) Timing and motor control in clumsy children. *J Mot Behav* 24:165–172.
- Wing A (1980) The long and short of timing in response sequences. In Stelmach GE, Requin J (eds), *Tutorials in Motor Behavior*. Amsterdam: North-Holland.
- Wing AM, Kristofferson AB (1973) The timing of inter-response intervals. *Percept Psychophys* 13:455–460.
- Woodrow H (1909) A quantitative study of rhythm. *Arch Psychol* 14:1–66.
- Woodrow H (1911) The role of pitch in rhythm. *Psychol Rev* 18:54–77.
- Woodrow H (1951) Time perception. In Stevens SS (ed), *Handbook of Experimental Psychology*. New York: John Wiley, pp. 1224–1236.
- Wundt W (1874) *Grudzüge der physiologischen psychologie*. Leipzig: Engelmann.

Chapter 7

Neurodynamics of Music

Edward W. Large

7.1 Introduction

Music is a high-level cognitive capacity, similar in many respects to language (Patel 2007). Like language, music is universal among humans, and musical systems vary among cultures and depend upon learning. But unlike language, music rarely makes reference to the external world. It consists of independent, that is, self-contained, patterns of sound, certain aspects of which are found universally among musical cultures. These two aspects – independence and universality – suggest that general principles of neural dynamics might underlie music perception and musical behavior. Such principles could provide a set of innate constraints that shape human musical behavior and enable children to acquire musical knowledge. This chapter outlines just such a set of principles, explaining key aspects of musical experience directly in terms of nervous system dynamics. At the outset, it may not be obvious that this is possible, but by the end of the chapter it should become clear that a great deal of evidence already supports this view. This chapter examines the evidence that links music perception and behavior to nervous system dynamics and attempts to tie together existing strands of research within a unified theoretical framework.

The basic idea has three parts. The first is that certain kinds of musical structures tap into fundamental modes of brain dynamics at precisely the right time scales to cause the nervous system to resonate to the musical patterns. Exposure to musical structures causes the formation of spatiotemporal patterns of activity on multiple temporal and spatial scales within the nervous system. The brain does not “solve” problems of missing fundamentals, it does not “compute” keys of melodic sequences, and it does not “infer” meters of rhythmic input. Rather, it *resonates* to music. The second part is that certain aspects of this process can be described with concepts that are already well-developed in neurodynamics, including oscillation of neural populations, rhythmic bursting, and neural synchrony. Dynamical analysis

E.W. Large (✉)

Center for Complex Systems & Brain Sciences, Florida Atlantic University,
Boca Raton, FL 33431, USA
e-mail: large@ccs.fau.edu

enables the description of the time-varying behavior of neural populations at the level of macroscopic variables. This approach also provides a means for moving between physiological and psychological levels of description, allowing a rather direct link between universal principles of neurodynamics and universal elements of music. The third and final part is that dynamic pattern formation corresponds directly to our experience of music. In other words, perceptions of pitch and timbre, feelings of stability and attraction, and experiences of pulse and meter arise as spatiotemporal patterns of nervous system activity. Section 7.2 introduces some of the relevant concepts from neurodynamics. The subsequent three sections consider, respectively, three essential and universal elements of music.

7.2 An Introduction to the Neurodynamics of Music

7.2.1 *Dynamical Systems in Neuroscience and Psychology*

Over the past several years, enormous progress has been made toward detailed understanding of nervous system dynamics, and mathematical models are now available that capture this behavior with considerable precision. Models of single neurons at the level of ion channels have now been available for more than 50 years (Hodgkin and Huxley 1952), and more recently the dynamical analysis of single-neuron models has explained and categorized the various kinds of behaviors observed in single neurons (Hoppensteadt and Izhikevich 1997; Izhikevich 2007). Starting in the 1960s and 1970s, analyses of small networks of neurons began to clarify the behavior of local neural populations (Wilson and Cowan 1973; Kuramoto 1975; e.g., Amari 1977). For example, Fig. 7.1a shows the connections between individual members of local excitatory and inhibitory subpopulations that are sufficient to sustain oscillation. Dynamical systems analyses have shown how such connectivity leads to the emergence of various types of dynamic behaviors from such a simple system; these include spiking, oscillation, bursting, and even more complex patterns as shown respectively in Fig. 7.1b (e.g., Crawford 1994; Strogatz 2000; Stefanescu and Jirsa 2008). Recently, with the aid of massive computing power, large-scale simulations have begun to investigate global interactions among local neural populations. In one large-scale simulation of thalamocortical dynamics based on models of various individual neuron types, realistic connectivity among local populations (derived from diffusion tensor imaging, see Fig. 7.1c) led to spontaneous emergence of global spatiotemporal patterns, including waves and rhythms, and functional connectivity on different scales (Izhikevich and Edelman 2008).

Because most cognitive functions are subserved by interactions among brain networks distributed over various subcortical and cortical areas, the studies described above have the potential to elucidate the neurodynamic underpinnings of cognition. It has even been argued that certain features of the complex dynamics observed in neural systems correlate well with key aspects of conscious experience

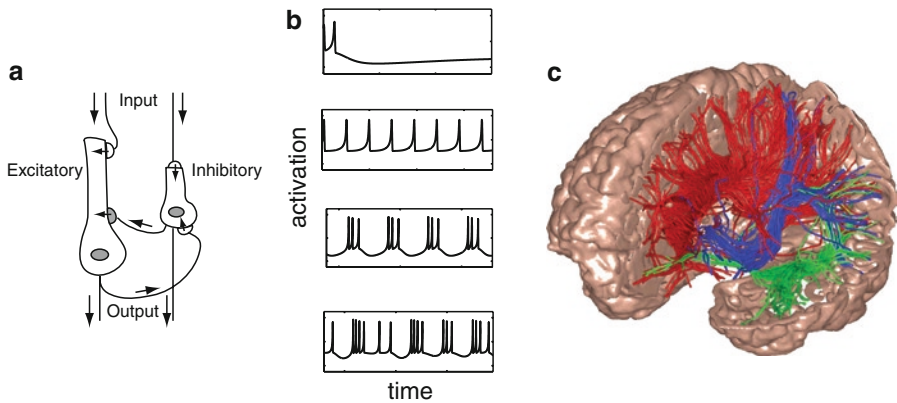


Fig. 7.1 (a) Neural oscillation can arise from interactions between excitatory and inhibitory neural subpopulations, shown visualized as one neuron representing each subpopulation. (Adapted from Hoppensteadt and Izhikevich, 1996a, with permission) (b) Time series illustrating different dynamical regimes for a single neuron within a local excitatory–inhibitory population of the type illustrated in A. Behaviors include spiking, oscillation, rhythmic bursting, and bursting intermixed with spiking. (From Stefanescu and Jirsa 2008, with permission). (c) Rendering of connections among local neural populations, obtained by means of diffusion tensor imaging data, as used in one large-scale dynamic thalamocortical simulation (From Izhikevich and Edelman 2008, with permission)

(e.g., Seth et al. 2006). Nevertheless, what we know about brain dynamics often seems to be disconnected from the observations we make and theories that we test at the level of human behavior (e.g., Barrett 2009). For example, linguistic theories of prosody, syntax, and semantics are not easily conceived in terms of the neurodynamics of synapses, neurons, and circuits (Poeppl and Embick 2005). Empirically, individual neurophysiological events are often observed to correlate with certain predictions about cognitive function (e.g., Kutas and Hillyard 1980; Tallon-Baudry and Bertrand 1999). However, behavior-level theories are generally not described at the level of neurodynamics (with some exceptions, e.g., Kelso 1995; Large and Jones 1999); rather, attempts are made to explain how brain dynamics might implement abstract computational mechanisms required by cognitive theories (see, e.g., Prince and Smolensky 1997; Jackendoff 2003).

Recent empirical and theoretical results suggest that, unlike linguistic communication, musical behavior may not require postulation of abstract computational mechanisms, but may be explainable directly in terms of neurodynamics. To facilitate understanding of this approach, this section introduces a few of the basic concepts of neurodynamics. The first is the notion of a local population of excitatory and inhibitory neurons, as illustrated in Fig. 7.1a. Such populations can give rise to several behaviors, illustrated in Fig. 7.1b, the simplest and most well-understood of which are oscillation, bursting and resonance. Because each of these behaviors has psychological significance, the remainder of this section describes oscillation and resonance in some detail, while bursting is visited toward the end of the chapter.

7.2.2 *Dynamical Systems and Canonical Models*

Because there are many different mathematical models that can be used to describe neural behavior, the principal concern is to choose a level of mathematical abstraction that is appropriate for the type of data that are available and the type of predictions that are desired. At the physiological level, individual neuron dynamics can be modeled in detail by Hodgkin–Huxley equations (Hodgkin and Huxley 1952), and more specialized models of neural oscillation are also available (FitzHugh 1961; Nagumo et al. 1962; Wilson and Cowan 1973; Hindmarsh and Rose 1984). It is important to keep in mind that actual neurons and neural networks are real dynamical systems. At any given time, a dynamical system has a state that can be modeled as a point in an appropriate *state space*. A dynamical model is a mathematical formalization – often a differential equation – that describes the time evolution of the point’s position in state space. *Stability* is a fundamental property of a dynamical system, which means that the qualitative behavior of its time evolution is not affected by small perturbations of the trajectory. Figure 7.1b shows four types of stable trajectories in excitatory–inhibitory neural networks. Section 7.2.3 discusses two important stable states: resting states (equilibria) and periodic trajectories (limit cycles). An attractor is a stable state to which a dynamical system evolves after a sufficiently long time. Thus, points that are close enough to an attractor return to the attractor even if temporarily disturbed, for example, by an external stimulus.

Returning to neurons and neural networks, resting states correspond to stable equilibria, and tonic spiking states correspond to limit cycle attractors (Izhikevich 2007). Analysis of the transition between states in dynamical models is called bifurcation analysis (Wiggins 1990). Bifurcation analysis is facilitated by the transformation of a complex dynamic model to a generic form, called a normal form. Interestingly, this analysis transforms virtually any model of neural oscillation into the same normal form, under certain assumptions that are generally reasonable for neural systems. This analysis reveals that neural oscillations share a set of universal properties, independent of many details (Wiggins 1990; Hoppensteadt and Izhikevich 1997). A *canonical model* is the simplest (in analytical terms) of a class of equivalent dynamical models, and can be derived using normal form theory. The canonical model we introduce in Eq. (7.3) was derived, using normal form theory, from a model of the interaction between excitatory and inhibitory neural populations (Wilson and Cowan 1973; Large et al. 2010). However, it is generic, so it could also be derived from other models of nonlinear oscillation (including outer hair cell models; see Julicher 2001). The canonical model uncovers universal properties, making predictions that hold under a rather general set of assumptions (Hoppensteadt and Izhikevich 1997). This makes the canonical model especially attractive from the point of view of modeling human perception and behavior. Some relevant generic properties of neural oscillation are described in Sect. 7.2.4.

Section 7.2.3 describes how the nervous system can resonate to sound, at various frequencies and on multiple timescales. The conceptual model is a network of

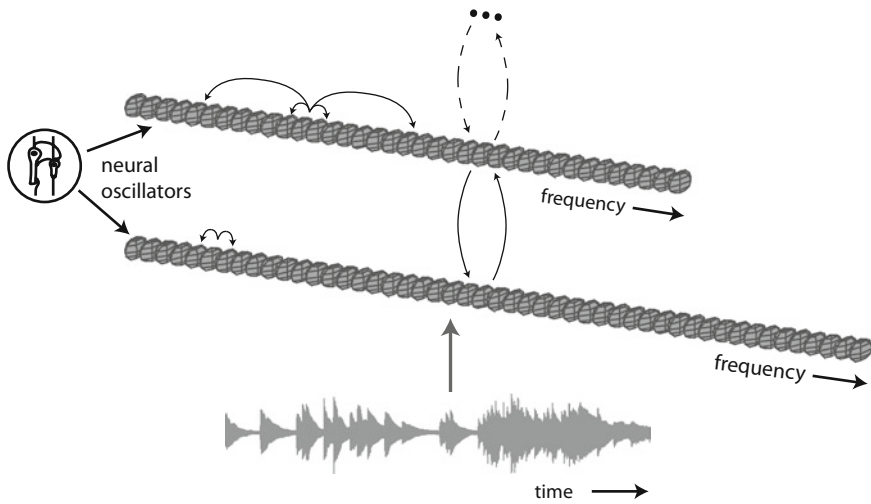


Fig. 7.2 Illustration of a layered neural architecture for processing acoustic stimuli. Each network layer consists of neural oscillators, arranged along a frequency gradient, from lowest to highest frequency. For pitch and melody, the first layer models cochlea, where connectivity between neighboring frequencies is shown. A second layer (e.g., dorsal cochlear nucleus) receives afferent stimulation from the first layer and also provides efferent feedback. Additional layers are possible, modeling neurons that phase lock action potentials to sound in higher auditory areas. Phase-locking to higher frequencies deteriorates as the auditory system is ascended, illustrated here as a lack of oscillators corresponding to higher frequencies in the second layer. Multilayer oscillator networks, operating at slower time scales, also serve as models for rhythm perception (see, e.g., Large 2000), and multilayer models could capture interactions between auditory and motor areas. Within the central nervous system connections between oscillators with different natural frequencies can be learned

oscillators, spanning a range of natural frequencies, stimulated with sound (Large et al. 2010). The basic idea is similar to signal processing by a bank of linear filters, but with the important difference that the processing units are *nonlinear*, rather than *linear* resonators. Such networks can be arranged into processing layers, as illustrated in Fig. 7.2. In what follows, this idea is applied to explain nonlinear resonance in the cochlea, phase-locked responses of auditory neurons, and entrainment of rhythmic responses in distributed cortical and subcortical areas.

7.2.3 Networks of Neural Oscillators Resonate to Sound

One way to understand nonlinear resonance is to first consider linear resonance. A common signal processing operation is frequency decomposition of a complex input signal, for example, by a Fourier transform. Often this operation is accomplished

via a bank of linear bandpass filters processing an input signal, $x(t)$. For example, a widely used model of the cochlea is a gammatone filter bank (Patterson et al. 1992), which – for comparison with our model – can be written as a differential equation:

$$\dot{z} = z(\alpha + i\omega) + x(t). \quad (7.1)$$

where the overdot denotes differentiation with respect to time (e.g., dz/dt), z is a complex-valued state variable, ω is radian frequency ($\omega = 2\pi f$, f in Hz), and $\alpha < 0$ is a linear damping parameter. The term $x(t)$ denotes linear forcing by a time-varying external signal. Because z is a complex number at every time, t , it can be rewritten in polar coordinates revealing system behavior in terms of amplitude, r , and phase, ϕ . This transformation is not reproduced here, but amplitude and phase of oscillations are discussed in Sect. 7.2.4. Resonance in a linear system means that the system oscillates at the frequency of stimulation, with amplitude and phase determined by system parameters. As stimulus frequency, ω_0 , approaches the oscillator frequency, ω , oscillator amplitude, r , increases, providing band-pass filtering behavior.

Recently, nonlinear models of the cochlea have been proposed to simulate the nonlinear responses of outer hair cells. It is important to note that outer hair cells are thought to be responsible for the cochlea's extreme sensitivity to soft sounds, excellent frequency selectivity, and amplitude compression (e.g., Egülüz et al. 2000). Models of nonlinear resonance that explain these properties have been based on the *Hopf normal form* for nonlinear oscillation, and are generic. Normal form (truncated) models have the form

$$\dot{z} = z(\alpha + i\omega + \beta |z|^2) + x(t) + h.o.t. \quad (7.2)$$

Note the surface similarities between this form and the linear resonator of Eq. (7.1). Again ω is radian frequency, and α is still a linear damping parameter. However, in this nonlinear formulation, α becomes a bifurcation parameter that can assume both positive and negative values, as well as $\alpha = 0$. The value $\alpha = 0$ is termed a bifurcation point and is discussed further in Sect. 7.2.4.1. $\beta < 0$ is a nonlinear damping parameter, which prevents amplitude from blowing up when $\alpha > 0$. Again, $x(t)$ denotes linear forcing by an external signal. The term *h.o.t.* denotes higher-order terms of the nonlinear expansion that are truncated (i.e., ignored) in normal form models. Like linear resonators, nonlinear oscillators come to resonate with the frequency of an auditory stimulus; consequently, they offer a sort of filtering behavior in that they respond maximally to stimuli near their own frequency. However, there are important differences in that nonlinear models address behaviors that linear ones do not, such as extreme sensitivity to weak signals, amplitude compression and high frequency selectivity; these are discussed in detail in the Sect. 7.2.4. The compressive gammachirp filterbank exhibits similar nonlinear behaviors, described within a signal processing framework (Irino and Patterson 2006; see also see Patterson et al., Chap. 2).

A canonical model was recently derived from a model of neural oscillation in excitatory and inhibitory neural populations (Wilson and Cowan 1973; Large et al. 2010). The canonical model (Eq. [7.3]) is related to the normal form (Eq. [7.2]; see e.g., Hoppensteadt and Izhikevich 1997), but it has properties beyond those of Hopf normal form models because the underlying, more realistic oscillator model is fully expanded, rather than truncated. The complete expansion of higher-order terms produces a model of the form

$$\dot{z} = z(\alpha + i\omega + (\beta_1 + i\delta_1)|z|^2 + \frac{(\beta_2 + i\delta_2)\varepsilon|z|^4}{1 - \varepsilon|z|^2}) + c\mathcal{P}(\varepsilon, x(t))\mathcal{A}(\varepsilon, \bar{z}) \quad (7.3)$$

There are again surface similarities with the previous models. The parameters, ω , α , and β_1 correspond to the parameters of the truncated model. β_2 is an additional amplitude compression parameter, and c represents strength of coupling to the external stimulus. Two frequency detuning parameters, δ_1 and δ_2 , are new in this formulation, and make oscillator frequency dependent on amplitude (see Fig. 7.4c). The parameter ε controls the amount of nonlinearity in the system. Most importantly, coupling to a stimulus is nonlinear (not discussed in further detail here, but see Large et al. 2010) and has a passive part, $\mathcal{P}(\varepsilon, x(t))$ and an active part, $\mathcal{A}(\varepsilon, \bar{z})$, producing nonlinear resonances that are discussed in Sect. 7.2.4.4. Helmholtz's (1863; see Sect. 7.3) difference tone, proposed to explain the pitch of the missing fundamental, was a passive nonlinearity. The three-frequency resonance of Cartwright et al. (1999a; see Sect. 7.3), proposed to explain residue pitch shift (Schouten et al. 1962), arises through the interaction between passive and active nonlinearities in this system. The canonical model given by Eq. (7.3) is more general than the Hopf normal form and encompasses a wide variety of behaviors that are not observed in linear resonators, some of which are discussed next.

7.2.4 Some Universal Properties of Nonlinear Oscillation

7.2.4.1 Andronov–Hopf Bifurcation

In the absence of stimulation, a nonlinear oscillator can display two qualitatively different stable states, both of which depend upon the specific value of the bifurcation parameter, α . Figure 7.3a illustrates the transition between a stable equilibrium and a stable limit cycle, called the Andronov–Hopf bifurcation. When $\alpha < 0$ the system behaves as a damped oscillator, but when $\alpha > 0$ (negative damping) the system generates a spontaneous oscillation. $\alpha = 0$ is the bifurcation point – also referred to as the critical value of the parameter – the value at which behavior changes from damped to spontaneous oscillation or vice versa. Other kinds of bifurcations that also lead to spontaneous oscillation can be found in this canonical model (see Guckenheimer and Kuznetsov 2007). Models of neural oscillation often assume spontaneous activity, i.e., $\alpha > 0$. Models of cochlear outer hair cells assume critical oscillation, i.e., $\alpha = 0$.

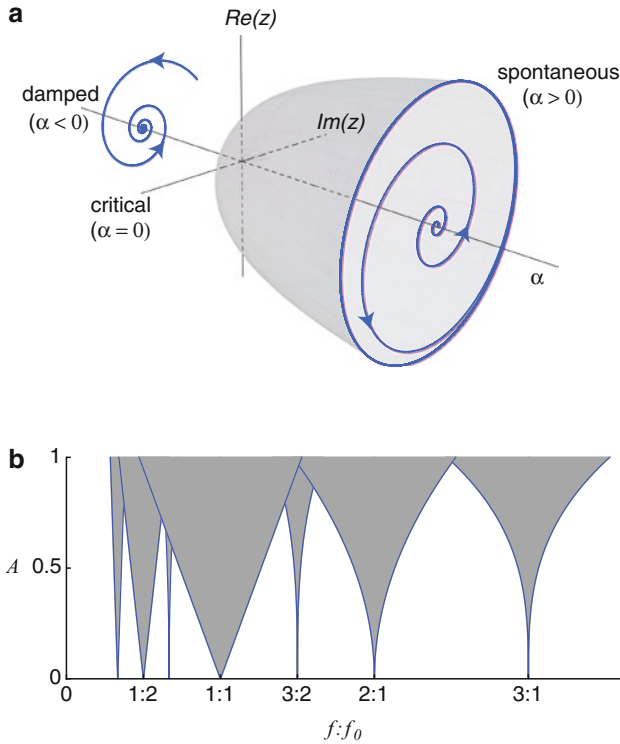


Fig. 7.3 (a) Andronov–Hopf bifurcation. The bifurcation diagram shows the two dimensions of the state space (the real and imaginary parts of z) and the value of the bifurcation parameter, α . When $\alpha < 0$ the oscillator displays damped oscillation. It responds passively to stimulus, then comes to rest (denoted by the inward spiral) at its stable fixed point. $\alpha = 0$ is referred to as the bifurcation point. At $\alpha = 0$ the system is poised exactly at the boundary between damped and spontaneous oscillation, a parameter regime called critical oscillation. Dynamical cochlear models assume critical oscillations of outer hair cells. When the bifurcation parameter becomes positive, the fixed point (rest) loses stability and limit cycle oscillation (denoted by the outward spiral) becomes the stable state. The system does not require a stimulus to sustain an active oscillation, but may phase lock to a stimulus if one is present. (b) An “Arnold tongues” bifurcation diagram showing some phase-locked regions for an active oscillator network. A denotes the stimulus frequency, and $f:f_0$ denotes the ratio of oscillator frequency to stimulus frequency, respectively. Nonlinear oscillators can respond to sinusoidal stimuli at near-integer ratio related frequencies, such as 1:2, 1:1, 3:2, and 3:1 (see also Fig. 7.4b and c). For active oscillation, there are well-defined boundaries between phase-locked (shaded areas) and non-phase-locked states

7.2.4.2 Entrainment

When the system oscillates spontaneously ($\alpha > 0$) and a stimulus is present, the oscillation will phase-lock, or *entrain*, to the stimulus. Figure 7.3b is a bifurcation diagram showing some phase-locked regions for an active oscillator network. Phase-locked states (Fig. 7.3b), are found at higher-order resonances (e.g., integer

ratios, discussed in Sect. 7.2.4.4) and are *stable*. On the horizontal axis, $f:f_0$ denotes the ratio of oscillator frequency to stimulus frequency, and A denotes the stimulus amplitude. The diagram shows regions of *attraction*, where an oscillator will adopt an instantaneous frequency that is different from its natural frequency. These oscillations are attracted to integer ratios of the stimulus frequency, such as 1:2, 1:1, 3:2, and 3:1 (see also Fig. 7.4b and c). For active oscillation, well-defined boundaries are found between phase-locked (shaded areas) and non-phase-locked states.

7.2.4.3 Nonlinear Amplitude Responses

Figure 7.4a illustrates the response of three different resonator models to sinusoidal stimulation, presented at their own natural frequencies. The curves show the amplitude responses for a linear filter (Eq. [7.1]), and two versions of a critical (i.e., $\alpha = 0$) nonlinear resonator, namely the Hopf normal form (Eq. [7.2]) and the fully expanded canonical model (Eq. 7.3). Linear filters have linear amplitude response. By contrast, both the Hopf normal form (truncated) model and the fully expanded canonical model exhibit extreme sensitivity to weak signals, one of the characteristic properties thought to explain nonlinear cochlear responses (e.g., Eguluz et al. 2000), discussed in Sect. 7.3. While both also exhibit amplitude compression, amplitude is fully compressive in the canonical model, but not in the Hopf normal form.

7.2.4.4 Higher-Order Resonance

Figure 7.4b shows the response of the three different resonator networks to a complex tone comprising two frequency components (f_1, f_2). Resonances are shown for a linear filter bank (Eq. [7.1]), and two versions of a critical oscillator array (i.e., $\alpha = 0$ for all oscillators), namely the Hopf normal form (Eq. [7.2]) and canonical model (Eq. [7.3]). Higher-order resonances are found only in the canonical network, due to the nonlinear coupling. Higher-order resonance means that a nonlinear oscillator network responds to a pure tone at the frequency f , with activity not only at f but also at harmonics ($2f, 3f, \dots$), subharmonics ($f/2, f/3, \dots$) and integer ratios ($2f/3, 3f/4, \dots$) of f . Further, if a complex tone is presented that contains multiple frequencies, a nonlinear network will respond at combination frequencies ($f_2 - f_1, 2f_1 - f_2, \dots$) as well. These responses follow orderly relationships and can be predicted given stimulus amplitudes, frequencies, and phases.

This feature of nonlinear resonance has important implications for understanding the behavior of such systems. A nonlinear oscillator network does not merely transduce signals; it actually adds frequency information, which may account for pattern recognition and pattern completion, among other things. The cochlea is known to produce audible higher-order resonances, including difference tones and harmonics (e.g., Robles et al. 1997), as produced by the canonical model. Neural pattern

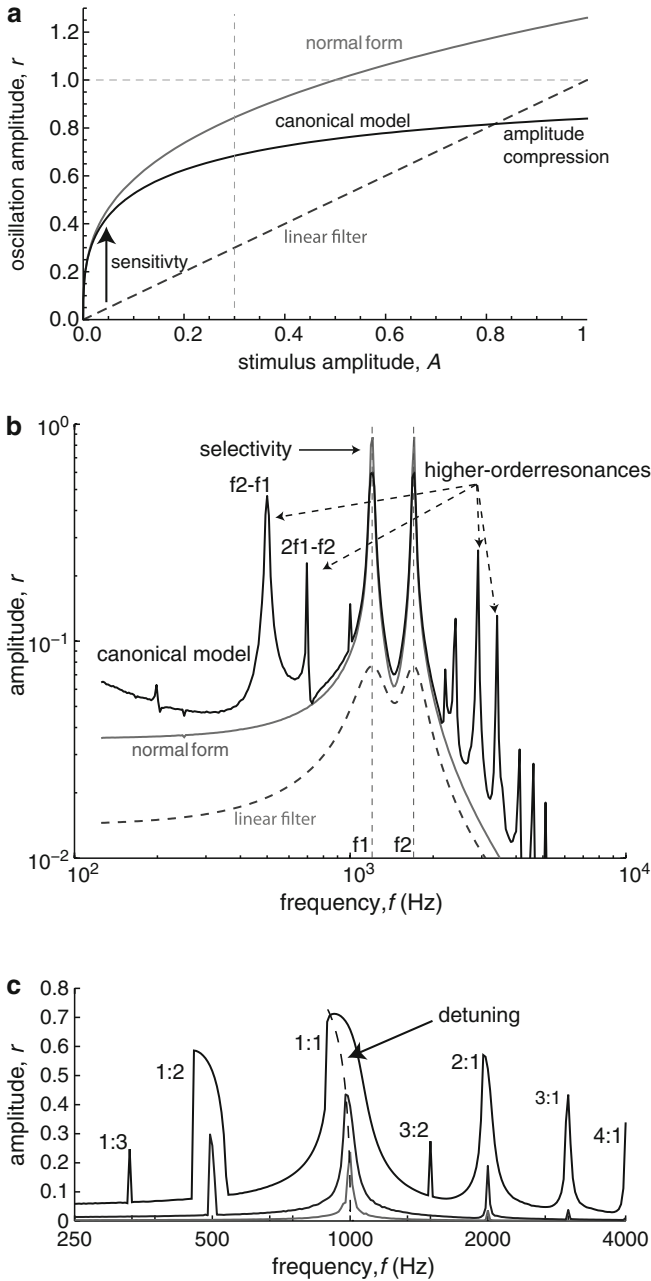


Fig. 7.4 Amplitude responses predicted by different resonance models. (a) A linear filter bank model (dashed line; Eq. [7.1]) vs. a critical Hopf normal form model (gray solid line; Eq. [7.2]) vs. a critical canonical model (solid black line; Eq. [7.3]) responding to stimuli at their own natural

completion based on nonlinear resonance may explain the perception of pitch in missing fundamental stimuli (Cartwright et al. 1999a), the perception of tonal relationships (e.g., Large and Tretakis 2005; Shapira Lots and Stone 2008), and the perception of pulse and meter in rhythmic patterns (Sect. 7.5; for a review, see McAuley, Chap. 6).

7.2.4.5 Frequency Selectivity and Detuning

Figure 7.4c presents the results of three simulations of an array of critical ($\alpha = 0$) nonlinear oscillators, based on Eq. (7.3). The frequencies of the oscillators in the array vary from 250 to 4,000 Hz, along a logarithmic frequency gradient, and the stimulus is a sinusoid with a frequency of 1,000 Hz. Each simulation shows the result for a different stimulus amplitude. These simulations illustrate two important properties of nonlinear resonance. First, the response at low stimulus amplitude levels reveals that high-frequency selectivity is achieved. As stimulus amplitude increases, frequency selectivity deteriorates due to nonlinear compression ($\beta_1, \beta_2 < 0$). Second, due to frequency detuning ($\delta_1, \delta_2 \neq 0$) the peaks in the resonance curve begin to bend as oscillator amplitude (r) increases. Both types of response agree with measurements in living intact cochleae (e.g., Ruggero 1992; see Fig. 7.5a). Also, as the stimulus amplitude increases, higher-order resonances appear at harmonics, subharmonics, and integer ratios.

7.2.4.6 Connectivity and Learning

Connections between oscillators can be modified, for example, via Hebbian learning (Hoppensteadt and Izhikevich 1996b), providing a mechanism for synaptic plasticity wherein the repeated and persistent coactivation of a presynaptic cell and a postsynaptic cell lead to an increase in synaptic efficacy between them. The number of possible synapses between excitatory and inhibitory subpopulations implies that a connection between two oscillators has both a strength and a natural phase (Hoppensteadt and Izhikevich 1996a). Both connection strength and phase can be learned by the Hebbian mechanism if a near-resonant relationship exists between their frequencies

←
Fig. 7.4 (continued) frequency. Amplitude response is linear for the linear filter, partially saturates for the normal form, and fully saturates for the canonical model. **(b)** Three resonator networks (linear, Hopf, and canonical) responding a two-frequency stimulus (f_1 and f_2). Oscillator amplitude is shown in logarithmic units as a function of resonator frequency. The canonical network produces harmonics and combination tones of the stimulus frequencies unlike the linear filter or the normal form model. **(c)** A canonical network (Eq. [7.3]) stimulated with a sinusoid at 1,000 Hz, for three different stimulus amplitudes (different curves). As stimulus amplitude increases, frequency selectivity deteriorates, frequency detuning is observed, and higher-order resonances appear

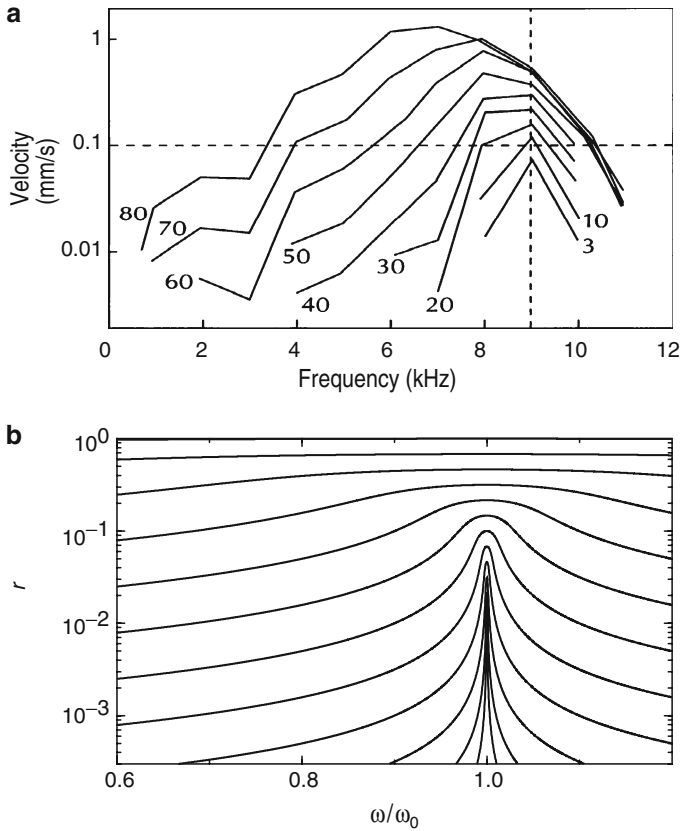


Fig. 7.5 (a) Laser velocimetric data from a living chinchilla's cochlea displaying the root-mean-square velocity of one point on the basilar membrane as a function of stimulus frequency. Each curve represents a different level of stimulation (dB SPL). Note the dramatic increase in bandwidth and the detuning as intensity increases. (From Ruggiero 1992, with permission) (b) Hopf resonance. The amplitude response, r , to different levels of forcing is obtained from Eq. (7.2); the amplitude of forcing increases in increments of 10 dB for successive curves from bottom to top. At resonance the response increases as the one-third power of the forcing, whereas away from the resonance the response is linear in the forcing (From Eguiluz et al. 2000, with permission)

(Hoppensteadt and Izhikevich 1996b). The Hebbian learning mechanism can learn connections between oscillators of different frequencies (Large *in press*).

7.2.5 Summary

Neural resonance can arise from the interaction between excitatory and inhibitory subpopulations. Canonical models of neural oscillation capture universal properties that are independent of physiological details. The same generic properties are also

found in other kinds of nonlinear oscillations, such as mechanical oscillations at the cellular scale (Choe et al. 1998; Eguíluz et al. 2000; Jülicher 2001). Canonical models are also available for burst oscillation, and these share some of the basic properties of limit cycle oscillation described in the preceding text (Izhikevich 2007). Although a detailed discussion of mathematical models of burst oscillation is beyond the scope of this chapter, the potential role of burst oscillation in rhythm perception is considered toward the end of the chapter. Gradient frequency networks of nonlinear oscillators can resonate to sound. Nonlinear resonators share some filtering properties with linear resonators, but also exhibit many properties that are not found in linear resonators. These include spontaneous oscillation, nonlinear amplitude responses, and higher-order resonance. Higher-order resonance is of critical importance; it implies a sort of pattern-formation behavior that is appropriate for describing the perception of structured patterns in musical sounds. For neural oscillation, there is also a canonical version of the Hebbian learning rule, enabling the development of connectivity among neural oscillators. The following sections will consider cochlear resonance, central auditory nonlinearities and entrainment of cortical rhythms from a dynamical systems point of view. The dynamic approach will lead to an understanding of the relationship between such phenomena and experiences of pitch, tonality, and rhythm in music perception.

7.3 Cochlear Resonance, Neural Resonance, and Pitch Perception

The first attempts to explain the physical basis of music perception concerned pitch. Shortly after Fourier methods were developed, Ohm (1843) proposed that pitch was a consequence of the auditory system's ability to perform Fourier analysis on acoustical signals. In Ohm's view, the pitch of a complex tone was a Fourier component of the sound. Helmholtz (1863) agreed that the ear acts as a rough Fourier analyzer and proposed the hypothesis that the analysis was performed by the basilar membrane. He described the cochlea as a time-frequency analysis mechanism that decomposes sounds into sinusoidal components for subsequent analysis by the central auditory nervous system. In the 1960s von Békésy (1960) demonstrated experimentally that the hypothesis of Helmholtz was essentially correct, that is, the basilar membrane carries out a frequency analysis of acoustic stimuli. Von Békésy's observations – using measurements on human cadavers – suggested that cochlear responses are linear over the range of physiologically relevant sound intensities.

Since then, however, a number of problems have arisen with the notion that the cochlea performs a passive, linear analysis. The weakest audible sounds impart energy per cycle no greater than that of thermal noise (Bailek 1987), and the system operates over a range of intensities that span at least 14 orders of magnitude. Gold (1948) recognized that these properties were incompatible with a passive, linear cochlea; rather, additional energy must be added into the system by active feedback. He also noted that if an active resonator underwent a Hopf bifurcation (see Fig. 7.4a),

it would oscillate spontaneously, and the ear would emit sound. Recently, the discovery of spontaneous otoacoustic emissions (Kemp 1979; Murphy et al. 1996) confirmed Gold's prediction. Moreover, laser-interferometric velocimetry performed on living, intact cochleae has revealed exquisitely sharp mechanical frequency tuning, which deteriorates with increasing stimulus amplitude (Ruggero 1992; Ruggero et al. 1997), as illustrated by different curves in Fig. 7.5a. These and related discoveries have led to the proposal that active amplification, in the form of Andronov–Hopf type nonlinearities, is the basic mechanism of the mammalian cochlear response (Choe et al. 1998; Camalet et al. 1999). The sharp mechanical frequency tuning, exquisite sensitivity, and operating range of the cochlea are now explained as self-tuned critical oscillations of hair cells (Eguíluz et al. 2000; Fig. 7.5b). It appears that the cochlea performs a type of active, nonlinear time-frequency transformation, using a network of locally coupled outer hair cell oscillators, each tuned to a distinct intrinsic frequency (eigenfrequency), and driven by an external stimulus (Duke and Julicher 2003; Kern and Stoop 2003; see also Irino and Patterson 2006).

Regarding perception, Seebeck (1841) demonstrated that if most of the energy at the fundamental frequency is removed from the complex spectrum of a periodic sound, the perceived pitch remains unchanged, matching the pitch of a sinusoid with the frequency of the missing fundamental.¹ Seebeck (1843) proposed a periodicity detection theory for pitch perception in complex sounds. However, Helmholtz (1863) embraced Ohm's approach, proposing that a physical component at the missing fundamental frequency, a "difference combination tone," could be generated by passive nonlinearities of the ear (similar to that in Eq. [7.3]). But Schouten et al.'s (1962) famous pitch-shift experiments demonstrated that the missing fundamental is not a difference tone. Schouten's theory of pitch was based on the periodicity properties of the nonresolved "residue" components of the stimulus. Eventually, because peripheral theories failed to explain psychophysical experiments and because dichotically presented stimuli also elicit pitch perception (e.g., Houtsma and Goldstein 1972), central processor theories for pitch perception arose (e.g., Goldstein 1973; Terhardt 1974). Complex pitch perception is still debated by theorists. It is determined neither solely by the spectral content of sound nor solely by its temporal structure (Plack and Oxenham 2005).

Recently, key theoretical advances have been made in understanding multifrequency resonance behaviors of nonlinear oscillators (Cartwright et al. 1999b), and this may have relevance for auditory perception. In experiments and numerical simulations, Cartwright and colleagues worked out the organization of higher-order resonances in representative nonlinear oscillators, and argued convincingly that such organization is universal across a large class of systems. They further showed that nonlinear resonance explains the "pitch shift of the residue," one of the important unexplained cases of pitch perception (Schouten et al. 1962; Cartwright et al. 1999a). If pitch depends on a difference tone (a passive nonlinearity), then when the

¹ Schouten (1938) showed that removing the fundamental component completely from the acoustic stimulus did not alter the pitch, and Licklider (1956) showed that the same pitch was heard even when the frequency region that would normally be occupied by the fundamental was masked by noise.

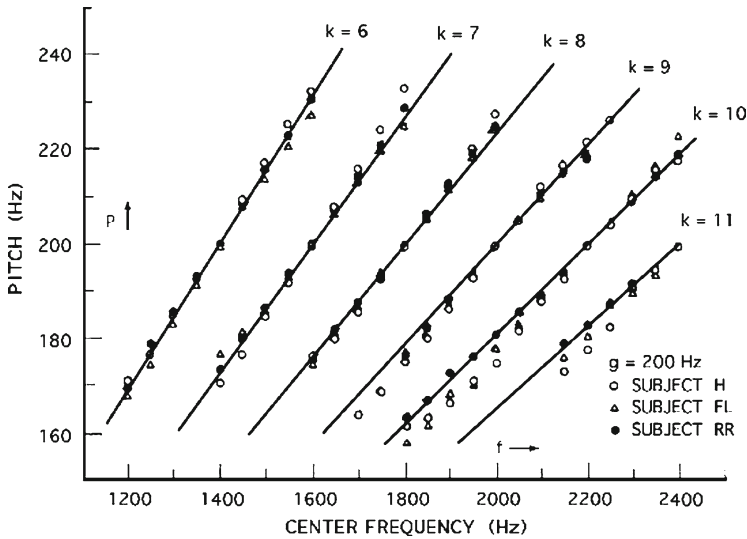


Fig. 7.6 Plot of the predicted (solid lines) pitch shift effect against the data of Schouten et al. (1962). Stimuli were tone complexes created from three successive harmonics of 200 Hz. Different lines correspond to different stimuli, and k is the harmonic number of the lowest frequency in the complex (e.g., $k = 6$ refers to harmonics 6, 7, and 8). Center frequency of the complex is plotted on the horizontal axis, and reported pitch on the vertical axis. Nonlinear resonance explains these data with considerable precision (From Cartwright et al., 1999a, with permission)

components of a missing fundamental harmonic complex are all shifted by the same amount the pitch should not change, because their difference remains the same. But Schouten et al. showed that perceived pitch does indeed shift, as illustrated in Fig. 7.6 for harmonics of 200 Hz. Figure 7.6 also shows that physical frequencies produced by generic nonlinear oscillators, acted upon by two independent periodic excitations, can reproduce the experimental data from Schouten's famous pitch-shift experiments with impressive precision. This provides strong evidence that nonlinear resonance is a viable neural mechanism for pitch perception. In Eq. (7.3), the pitch-shift resonance of Cartwright and colleagues arises through the interaction between passive and active nonlinearities and is nontrivial. Thus, higher-order resonance of neural oscillation could explain important aspects of pitch perception.

Nonlinear oscillations can arise through the interaction of excitatory and inhibitory neural populations, as illustrated in Fig. 7.1a and b, and there is a growing body of evidence consistent with nonlinear oscillation in the central auditory system. In mammals, action potentials phase-lock to both fine time structure and temporal envelope modulations at many different levels in the central auditory system, including cochlear nucleus, superior olive, inferior colliculus (IC), thalamus, and A1 (Langner 1992; Joris et al. 2004), and recent evidence points to a key role for synaptic inhibition in maintaining central temporal representations. Hyperpolarizing inhibition is phase-locked to the auditory stimulus and has been shown to adjust the temporal sensitivity of coincidence detector neurons (Grothe 2003), while stable pitch representation in the

IC may be the result of a synchronized inhibition originating from the ventral nucleus of the lateral lemniscus (Langner 2007). Such evidence suggests that nonlinear oscillation may be a good model for phase-locked central auditory responses.

Recent evidence also supports higher order resonance in neural activity. Multip peaked spectrotemporal receptive field (STRF) curves have been identified in cat primary auditory cortex, some with responses to second and third harmonics of the fundamental frequency (Sutter and Schreiner 1991). Modulation-rate selective cells in the auditory midbrain of *Pollimyrus*, which receive both excitatory and inhibitory input, have been successfully modeled as nonlinear oscillators (Large and Crawford 2002). Nonlinear STRFs have been identified in cat IC (Escabi and Schreiner 2002), and neurons in the IC of the gerbil have been observed to respond at harmonic ratios (e.g., 3:2, 2:1, 5:2; cf. Fig. 7.4c) with the temporal envelope of the stimulating waveform (Langner 2007). Nonlinear $2f_1 - f_2$ difference tones (see Fig. 7.4b) have been identified in brain stem auditory evoked potentials of guinea pigs (Chertoff and Hecox 1990), in human frequency-following responses using electroencephalography (EEG; Pandya and Krishnan 2004), and in auditory cortex, using steady-state methods in magnetoencephalography (MEG) (Purcell et al. 2007). These results provide evidence of higher-order resonance in the auditory system all the way from the cochlea to the primary auditory cortex.

7.3.1 Summary

The auditory nervous system is highly nonlinear, and observed responses are consistent with the generic predictions of nonlinear resonance, possibly arising in excitatory–inhibitory networks of the auditory nervous system. One potentially important functional consequence would be the perception of pitch, which may arise through an active nonlinear mechanism that is generic to nonlinear oscillators (i.e., Eq. [7.3]). Fourier-based approaches rely on linear systems theory almost exclusively, thus they describe human perceptual capabilities only approximately. However, generic models of neural oscillation (e.g., Eqs. [7.2] and [7.3]) are available, which are able to capture functionally important nonlinearities. As a result, such models may be able to capture many human perceptual and cognitive capabilities in a physiologically realistic way, but without strong dependence on physiological details. This observation has important implications not only for pitch perception, but also for other aspects of musical experience.

7.4 Neurodynamics of Tonality

The preceding section focused on responses to individual tones. But music is more than the perception of isolated tones; it involves the combination of tones into larger structures, such as melodies. Musical melodies typically involve discrete

tones, organized in archetypal patterns that are characteristic of musical genres, styles, and cultures. These patterns may be related to a *scale*, an ordered collection of all the tones used in a given melody, which summarizes the frequency ratios that govern the intervals between tones in a melody. One feature the melodies of most musical systems share is that they give rise to tonal percepts. Listeners experience feelings of *stability* and *attraction* among tones in a tonal melody. Stability means that one or more tones are perceived as points of repose. One specific tone, called the tonic, provides a focus around which the other tones are dynamically organized, and there is a hierarchy of relative stability, such that some tones are perceived as more stable than others. Less stable tones provide points of dissonance or tension; more stable tones provide points of consonance or relaxation. Less stable tones are heard relative to more stable ones, such that more stable tones are said to attract the less stable tones (e.g., Lerdahl 2001). Some theorists have described tonal attraction by analogy to physical forces, such as gravity (Larson 2004); others link it to the resolution of musical dissonance (Bharucha 1984). Zuckerkandl (1956) argued that these dynamic tonal qualities make “melodies out of successions of tones and music out of acoustical phenomena (p. 21).” But what processes in the nervous system could give rise to such perceptions in music?

The oldest theory of musical consonance is that perceptions of consonance and dissonance are governed by ratios of whole numbers. Pythagoras is thought to have first articulated the principle that intervals of small integer ratios (cf. Figs. 7.3b and 7.4c) are pleasing because they are mathematically pure (Burns 1999). He used this principle to explain the musical scale that was in use in the West at the time, and Pythagoras and his successors proposed small-integer-ratio systems for tuning musical instruments, such as Just Intonation (JI). Modern Western equal temperament (ET), divides the octave into 12 intervals that are precisely equal on a log scale. ET approximates JI, and transposition in ET is perfect, because the frequency ratio of each interval is invariant. Apart from octaves, however, the intervals are not small integer ratios, they are irrational. The fact that intervals based on irrational ratios are approximately as consonant as nearby small integer ratios is generally considered *prima facie* evidence against the theory that musical consonance derives from the mathematical purity of small integer ratios.

Helmholtz (1863) hypothesized that the dissonance of a pair of simultaneously sounding complex tones was due to the interference of its pure tone components, explaining dissonance as a sensation of roughness produced by the beating of sinusoids. This phenomenon, called sensory dissonance, is heard when simultaneous tones interact within an auditory critical band (Plomp and Levelt 1965), and the interaction of pure tone components correctly predicts ratings of consonance for pairs of complex tones (Kameoka and Kuriyagawa 1969). However, there are a number of problems that arise with Helmholtz theory as a theory of musical consonance (Dowling and Harwood 1986). For one thing, the sensory dissonance phenomenon is heard for isolated clusters of simultaneously sounded tones, but not for sequentially presented tones (i.e., melodies). Moreover, musical consonance and dissonance are intrinsically dynamic: “... a dissonance is that which requires resolution to a consonance” (Dowling and Harwood 1986). Recently, Shapira Lots and

Stone (2008) used the theory of coupled neural oscillators to explain why simple frequency ratios are important for music perception. They used the width of the resonance regions (cf. Fig. 7.3b) for higher order resonances to predict the consonance of the intervals of chromatic scales. Their analysis revealed that this method of ordering higher order resonances corresponds to the standard ordering of consonance often listed in Western music theory (Helmholtz 1863), suggesting that neural synchrony may be important in music perception. One piece of relevant evidence comes from a recent study in which nonlinear responses to harmonic musical intervals were measured in the auditory brain stem response. Two musical intervals, the major sixth (E3 and G2) and the minor seventh (E3 and F#2), were found to give rise to highly nonlinear responses including difference and summation tones (cf. Fig. 7.4b), revealing nonlinear processing of simultaneously sounded musical intervals in the auditory system (Lee et al. 2009).

Tonal perceptions such as stability, attraction, perceptual categorization, and learning of tonal relationship may depend on neural resonance as well (Large and Tretakis 2005; Large *in press*). Perceptual categorization and discrimination experiments reveal that musicians show categorical perception of melodic intervals (Burns and Campbell 1994), and nonmusicians also perceive pitch categories (Smith et al. 1994). Resonance regions (Fig. 7.3b) predict perceptual categorization of musical intervals, because resonances not only affect oscillators with precise integer ratios; they also establish patterns of resonant neighborhoods (Fig. 7.7c). Thus, even if resonance center frequencies do not precisely match stimulus frequencies, as connection strength increases, larger regions of the network resonate, emanating from integer ratios, and encompassing nearby ratios (Large and Tretakis 2005).

Simplicity of frequency ratios has been shown to account not only for judgments of consonance and dissonance, but also for judgments of similarity and discrimination of tone patterns across a wide range of tasks and listeners (Schellenberg and Trehub 1994). In one study, 6-month-old infants detected changes to sequentially presented pairs of pure tones (intervals) only when the tones were related by simple frequency ratios (Schellenberg and Trehub 1996). In adults as well, changes from patterns with simple frequency ratios to those with more complex ratios were more readily detected than were changes from complex ratios to simpler ratios. This implies that memories for tone sequences with small integer ratio relationships are more stable than memories for complex integer relationships. Large (*in press*) found a similar result in an oscillator network simulation. Tones with small integer ratio relationships (1:1, 5:4 and 3:2 – a tonic triad) produced a stable memory in the neural oscillator network (cf. Fig. 7.2). Although a leading tone (8:15 ratio with the tonic frequency) could be stabilized through external stimulation, when the external stimulus was removed, the leading tone frequency lost stability as those oscillators that had responded at the leading tone frequency began to resonate at the tonic frequency. In other words, the tonic frequency functioned as an attractor of nearby oscillations. Thus, nonlinear resonance predicts both memory stability of small integer ratios and tonal attraction among sequentially presented frequencies (Large *in press*).

Krumhansl and Kessler (1982) measured the stability of each tone within a musical key directly, by asking listeners to rate how well individual pitches fit within a

tonal context (see Fig. 7.7a, b). Higher goodness-of-fit ratings imply higher stability, so, for example, C and G are the most stable in both the C-major and C-minor tonal contexts. When applied to Western music, the measured hierarchies are consistent with music-theoretic accounts and agree with frequency-of-occurrence statistics for tonal songs (Krumhansl 1990). It is possible to apply a dynamic analysis to predict tonality rating data. Nonlinear resonance predicts that the relative stability of higher order resonances is given by $\varepsilon^{(k+m-2)/2}$, where k and m are the numerator and denominator, respectively, of the frequency ratio (Hoppensteadt and Izhikevich 1997). Here ε is a parameter that controls coupling nonlinearity (see Eq. [7.3]). One could use this fact and assume that tones heard in a tonal context would be stabilized in memory (based on the simulation results described above), to create a single-parameter (ε) fit to the stability judgments. Theoretical predictions of stability based on this analysis matched perceptual judgments well (Large in press), as shown in Fig. 7.7a and b.

The tuning systems of the world’s largest musical cultures, Western, Chinese, Indian, and Arab-Persian, are based on small integer ratio relationships (Burns 1999).² However, each tuning system is different, and this has led to the notion that

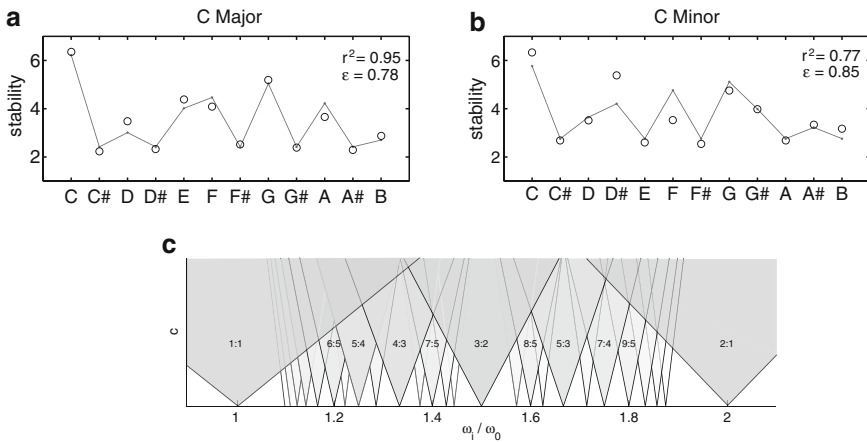


Fig. 7.7 Comparison of theoretical stability predictions and human judgments of perceived stability for two Western modes: (a) C-major and (b) C-minor. Open circles denote mean goodness-of-fit ratings from Krumhansl and Kessler (1982), and solid lines represent nonlinear resonance predictions. (c) An “Arnold tongues” bifurcation diagram showing natural resonances in a gradient frequency nonlinear oscillator array as a function of connection strength and frequency ratio (similar to Fig. 7.3b). An infinite number of resonances are possible on this interval; shown here are the unison (1:1), the octave (2:1), and the 25 most stable resonances in between. (From Large in press, with permission) Shading of each resonance region reflects the intrinsic stability of the ratio, used for the predictions shown in (a) and (b). Where regions overlap, less stable frequencies are attracted to more stable frequencies

²ET in the West is designed to approximate small integer ratio tuning and has been in widespread use for less than 150 years.

frequency relationships do not matter in high-level music cognition; rather, auditory transduction of musical notes results in abstract symbols, as in language (see, e.g., Patel 2007). If this were true, stability and attraction relationships would also have to be learned presumably based solely on the frequency-of-occurrence statistics of tonal music (for a current overview, see Krumhansl and Cuddy, Chap. 3). However, Hebbian learning of multifrequency relationships can provide a theoretical basis for the acquisition of frequency relationships. As the music of one's culture is heard, auditory networks would learn the most stable attractors whose center frequencies closely approximate the experienced relationships. Natural resonances predict significant constraints on which frequency relationships can be learned, as illustrated in Fig. 7.7c. Hebbian synaptic modification would effectively prune some resonances, while retaining or enhancing others (Large *in press*). This reasoning suggests that frequency relationships are learned depending on the frequency relationships employed in the music of a particular style or culture. However, stability and attraction relationships are not learned per se, but are intrinsic to neural dynamics given a particular set of frequency relationships.

7.4.1 Summary

Nonlinear resonance predicts the perceived dynamics of tonal organization and important aspects of neurophysiological responses, qualitatively and quantitatively. Thus, nonlinear resonance may provide the neural substrate for a substantive musical universal, similar to the concept of universal grammar in linguistics (Prince and Smolensky 1997). However, in the case of music, perceptual universals are predicted by universal properties of nonlinear resonance, properties that provide direct links to neurophysiology. Learning would alter connectivity to establish different resonances and different tonal relationships. According to this approach, stability and attraction relationships would not be learned on the basis of statistical properties of tone sequences; instead, because nonlinear resonance predicts stability and attraction, and because stability and attraction are correlated with sequence statistics, nonlinear resonance predicts tone frequency statistics. Thus, higher-order resonances may create resonant tonal fields in the central nervous system, and musical melodies may be perceived in relation to such fields, creating a dynamical context within which perception of tone sequences takes place.

7.5 Resonating to Rhythm

Musical structure is found not only in the pitch dimension, but also in the time dimension. Jones (1976) originally proposed that neural rhythms entrain to the temporal structure of environment stimuli. Entrainment of intrinsic neural dynamics would enable

dynamic attending, providing a basis for temporal expectancy and facilitating perception of events that occur at expected points in time. Musical rhythms are highly temporally structured sequences of acoustic events, and in most musical rhythms people perceive periodicity, called pulse or beat, and structured patterns of accentuation among pulses, called meter (London 2004). Pulse can be thought of as a frequency, and meter as a pattern of frequencies, which can be transcribed as arrangements of dots (reflecting beats) aligned with a musical score, as shown in Fig. 7.8a. The fundamental pulse periodicity (the rate at which one taps with a rhythm) is notated as a single row of beats, and the pattern of strong and weak pulses as additional rows of beats at related frequencies (for a more thorough discussion of pulse and meter, see McAuley, Chap. 6). Sometimes metrical frequencies are physically present in stimulus rhythms; sometimes they are not. For example, in the clave rhythm of Fig. 7.8a, the frequencies of the pulse and meter are almost completely absent.

The temporal relationships observed in human musical interactions are among the most elaborate observed in nature (for a review, see Large 2008). When humans

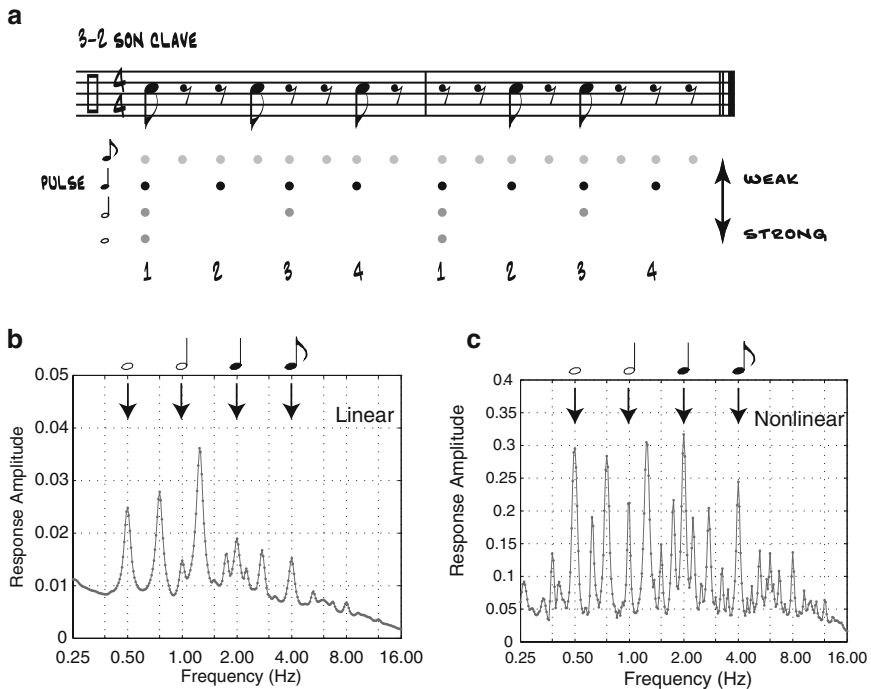


Fig. 7.8 (a) Pulse and meter of the 3–2 son clave rhythm à la Lerdahl and Jackendoff (1983). At 500 ms/quarter note, the pulse frequency would be 2 Hz or 120 bpm. Results of a linear (b; Eq. [7.1]) and nonlinear (c; Eq. [7.3]) analysis of the 3–2 son clave rhythm. The linear analysis reveals very little energy in this rhythm at the pulse frequency or at other metrical frequencies. The nonlinear analysis responds at all metrical frequencies (as well as many others) via higher-order resonance

temporally coordinate in musical interactions, we synchronize – or more generally, entrain – pulse frequencies. Entrainment is the process whereby two spontaneously oscillating systems, which have different frequencies when they function independently, assume the same frequency, or integer-ratio related frequencies, when they interact. In general, entrainment of neural oscillations predicts multifrequency coordination at simple frequency ratios such as 1:1, 1:2, 1:3, 2:3, due to higher-order resonance (Figs. 7.3b and 7.4c). Such entrainment is found in everyday musical interactions and has been observed in behavioral studies involving perception (e.g., Vos 1973), attention (e.g., Barnes and Jones 2000), and motor coordination (e.g., Parncutt 1994). Moreover, newborns can perceive pulse (Winkler et al. 2009); by 7 months infants discriminate rhythms and categorize melodies on the basis of meter (Hannon and Johnson 2005); and 9-month-old infants detect changes in the context of metric rhythms but not in sequences that induce a metric percept only weakly or not at all (Bergeson and Trehub 2006). Toddlers as young as 2.5 years are capable of entraining motor rhythms with periodic sequences (Provasi and Bobin-Begue 2003; Kirschner and Tomasello 2009), and even some animals can entrain motor rhythms to music (Patel et al. 2009; Schachner et al. 2009).

The two complexities of rhythm that are the most troublesome for theoretical accounts of pulse and meter are *syncopation* and *temporal fluctuation*. Syncopation refers to rhythms in which accented events occur on weaker positions in the metrical structure while leaving nearby stronger positions empty (Fitch and Rosenfeld 2007). This is illustrated by the clave rhythm of Fig. 7.8a, in which note events occur on only half the beats of the basic pulse and occur often on relatively weak beats. Temporal fluctuation refers both to localized temporal nuances and to larger scale tempo changes (i.e., rubato) that arise in music performance due to motoric, perceptual, and expressive constraints (Palmer 1997; Penel and Drake 1998). Temporal fluctuation is correlated with important aspects of musical structure (Sloboda 1983, 1985; Todd 1985; Palmer 1989), exhibits *1/f* (fractal) structure (Rankin et al. 2009), and conveys affect and emotion to listeners (Sloboda and Juslin 2001).

Several studies have compared people's ability to entrain to simply structured versus syncopated rhythms (Snyder and Krumhansl 2001; Toiviainen and Snyder 2003; Patel et al. 2005). Level of syncopation is a good predictor of pulse-finding difficulty, and syncopation causes some off-beat taps and some switches between on beat and off beat tapping (Snyder and Krumhansl 2001; Patel et al. 2005). Overall, however, humans are quite good at entraining to the pulse of even highly syncopated rhythms. How is this possible? Figure 7.8b and c illustrates two resonance predictions for a highly syncopated rhythm (Fig. 7.8a), one generated by a linear filter bank (Eq. [7.1]) and the other by a critical nonlinear resonator array (Eq. [7.3]). The linear filter bank responds at frequencies that are physically present in the time series, finding very little energy at 2 Hz (the pulse frequency for this rhythm). There are several strong peaks, however, with the strongest at 1.33 Hz, corresponding to the time interval between the first and second notes (i.e., 1/0.750 s). By contrast, a nonlinear oscillator array finds its strongest peak at 2 Hz, the pulse frequency, due to higher-order resonance. Such observations predict that perceived pulse in highly syncopated rhythms arises through higher order resonance.

Most listeners are also good at synchronizing with music that contains expressive timing fluctuations (Drake et al. 2000). A number of studies have investigated the response of nonlinear oscillators to temporal fluctuation in music (McAuley 1995; Toiviainen 1998; Large and Palmer 2002; Cont 2008), generally finding support for oscillator predictions. However, one surprising recent finding in this area is that people are able to predict, rather than simply react to, expressive temporal fluctuations (Repp 2002; Rankin et al. 2009). It has been hypothesized that listeners exploit musical and fractal structure to predict tempo changes in music (Repp 2002; Rankin et al. 2009). One important aspect of this ability may be the covert monitoring of multiple metrical frequencies during entrainment (Large et al. 2002; Repp 2008). One nonlinear resonance model captures this phenomenon as coupling between nonlinear oscillators as they respond at different metrical frequencies (Large and Jones 1999; Large and Palmer 2002; Jones 2008).

Recent functional imaging studies have shown that the perception of rhythmic sequences involves multiple, spatially distinct brain regions. Rhythmic information is represented across broad cortical and subcortical networks in a manner that is dependent upon task and rhythmic complexity (Sakai et al. 1999; Grahn and Brett 2007; Jantzen et al. 2007; Chen et al. 2008). Metric rhythms are easier to reproduce, and elicit higher activity in the basal ganglia and supplementary motor area (Grahn and Brett 2007), suggesting that these motor areas play a role in mediating pulse and meter perception. Both performance and neural activity are modulated as musicians and nonmusicians tap in synchrony with progressively more syncopated auditory rhythms (Chen et al. 2008). In perception, secondary motor regions were recruited in musicians and non-musicians, and the dorsal premotor cortex appeared to mediate auditory–motor interactions (Chen et al. 2008). The dorsal auditory pathway is also implicated in rhythm performance, regardless of the modality in which the rhythms are trained and paced (Karabanov et al. 2009). Thus, both auditory and motor areas play key roles in both rhythm perception and rhythm production. A set of brain areas including dorsal auditory pathway areas, dorsal premotor cortex, the supplementary and presupplementary premotor areas, the cerebellum, and the basal ganglia are implicated. A key question is: What is happening in this distributed network?

Using EEG, Snyder and Large (2005) observed that peaks in the power of induced beta- and gamma-band activity anticipated tone onset (average ~0 ms latency), were sensitive to intensity accents, and persisted when expected tones were omitted, as if an event had appeared. By contrast, evoked activity occurred in response to tone onsets (~50 ms latency) and was strongly diminished during tone omissions. Recent MEG studies have found subharmonic rhythmic responses in the beta-band when subjects were instructed to impose a subjective meter on a periodic stimulus (Iversen et al. 2009), and anticipatory responses for periodic and metrical sequences, but not for randomly timed sequences in primary auditory cortex (Fujioka et al. 2009). Thus, the features of high-frequency brain activity match the main predictions for pulse and meter. Such observations could indicate cortical bursting, which can also arise from excitatory–inhibitory neural circuits (Izhikevich 2007; see Fig. 7.1a). Bursting is a dynamic state where neurons repeatedly fire

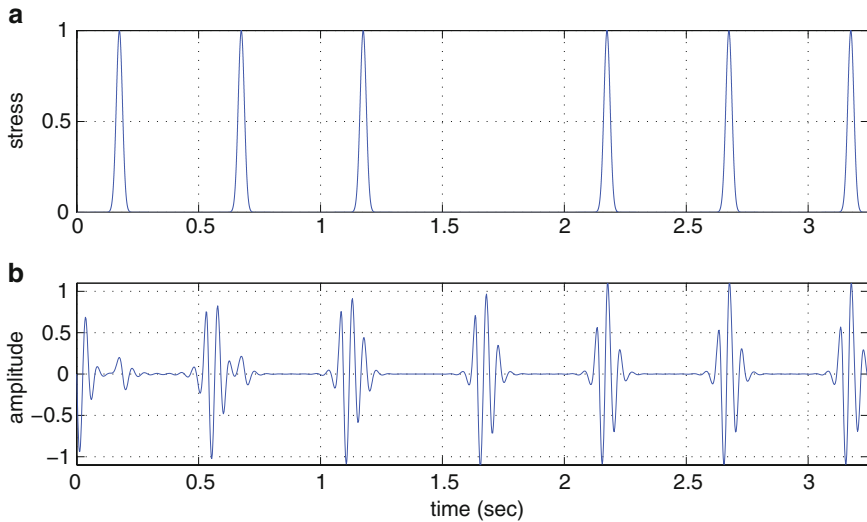


Fig. 7.9 Response of a burst oscillator (Izhikevich 2000) to a rhythmic pattern. **(a)** Continuous time series representation of event onsets. **(b)** Bursts of activity entrain to the stimulus and are observed even in the absence of a stimulus event (From Large 2008, with permission)

groups, or bursts, of action potentials, and each burst is followed by a period of quiescence before the next occurs (Izhikevich 2007). Interburst periods, the time interval between one burst and the next, are generally consistent with timescales of musical pulse and meter. Burst oscillation is currently receiving a great deal of attention in the computational neuroscience literature, and mathematical analyses have shown that rhythmic bursting displays key properties (Coombes and Bressloff 2005; Izhikevich 2007) that are necessary to predict pulse and meter. Figure 7.9 shows a computational simulation of burst oscillation (Izhikevich 2000) responding to a simple rhythm, displaying both entrainment to the stimulus sequence (Fig. 7.9a) and oscillatory persistence in the absence of an element in this sequence (Fig. 7.9b).

Moreover, bursts of high-frequency activity could explain communication between different cortical areas (Brovelli et al. 2004). For example, oscillatory activity in the beta range is widely observed in sensorimotor cortex in connection with motor behavior in humans (Pfurtscheller and Lopes da Silva 1999; Salenius and Hari 2003) and nonhuman primates (Rougeul et al. 1979; Sanes and Donoghue 1993; MacKay and Mendonca 1995). Synchrony of beta oscillations is often observed between different areas of sensorimotor cortex (Murthy and Fetz 1992; Sanes and Donoghue 1993). Moreover, synchronized beta oscillations may bind multiple sensorimotor areas into a large-scale network during motor behavior and carry causal influences from primary somatosensory and inferior–posterior parietal cortices to motor cortex (Brovelli et al. 2004). Anticipatory rhythmic bursts of beta activity may enable communication between auditory and motor cortices in rhythm perception and motor coordination as well. Rhythmic bursts of higher frequency gamma activity may also enable functional communication between different cortical regions. The theoretical picture that

emerges is one of communication, through bursts of high-frequency activity, between different neural areas as they resonate to rhythmic patterns.

Entrainment of rhythmic neural bursting could explain how the perception of pulse and meter arise from listening to complex sequences, as well as the development of expectancy for events occurring in a rhythmic context. Dynamic attending theory (DAT) hypothesizes that endogenous attentional rhythms entrain to temporally structured external events (Jones 1976; Large and Jones 1999). DAT has traditionally been discussed in terms of facilitation of perception to certain external events, and this has found support in a number of recent studies (McAuley and Kidd 1995; Jones and Yee 1997; Large and Jones 1999; Barnes and Jones 2000; Jones et al. 2002; Jones and McAuley 2005; Quené and Port 2005). However, conceptualizing attentional rhythms as rhythmic bursting provides a new hypothesis regarding the role of attention in coordinating the interaction between auditory and motor areas (Large and Snyder 2009). Bursts of beta and gamma band activity that entrain to external rhythms could provide a mechanism for rhythmic communication between distinct brain areas, and attention may facilitate such integration among auditory and motor areas.

7.5.1 *Summary*

Entrainment of endogenous neural rhythms and higher order resonance could explain why metrical percepts favor small integer ratios. It can also explain how people perceive a regular pulse in highly syncopated rhythms and how listeners adapt to frequency fluctuations in expressive performances. Rhythmic bursting in higher frequency bands is a plausible neural correlate of pulse and meter. This could explain not only perceptual facilitation of expected events, but also functional integration of auditory and motor areas.

7.6 **Summary and Conclusions**

As noted at the outset, it is informative to compare theories of music with theories of language. Poeppel and Embick (2005) discuss a “conceptual granularity mismatch” between cognitive and neurobiological mechanisms in language. That is, theories that are typically invoked to account for linguistic computation – in terms of syntax, meter, and semantics – are not related in any obvious way to the neurodynamics of synapses, neurons, and circuits. The theoretical picture they paint is potentially bleak and would seem to require a paradigm shift to reconcile the two approaches to language. However, it may be unnecessary to invite such theoretical difficulties into the musical domain. In music, our experiences of the fundamental universals, including pitch, tonality, and rhythm, can be readily conceived in relation to neurodynamic universals, including limit cycle oscillation, resonance, and rhythmic bursting.

Helmholtz (1863) originally envisioned that a proper understanding of auditory physiology should one day form the basis for a theory of music. However, the auditory system is highly nonlinear, and Poincaré, the father of modern dynamical systems theory, was only a boy when Helmholtz penned the preface to the first edition of *On the Sensations of Tone*. Modern theories of auditory and music perception were built on the foundation of linear resonance. Where linear resonance has proven insufficient to explain cognitive and perceptual phenomena, complex mechanisms and general purpose computation have been recruited to fill the explanatory gaps. Known auditory nonlinearities can be described with well developed concepts of modern neurodynamics. These phenomena are summarized in high-level dynamical models, called canonical models, which are appropriate for describing the macroscopic dynamics of neural populations and for describing key aspects of perception, cognition, and behavior. Neurodynamic models seem to capture many features of music perception and behavior in their own terms, without the need to resort to more abstract computational descriptions. These observations suggest that our qualitative experiences of music arise as a direct consequence of the interaction of sound with the intrinsic dynamics of the nervous system.

References

- Amari S (1977) Dynamics of pattern formation in lateral inhibition type neural fields. *Biol Cybern* 27:77–87.
- Bailek W (1987) Physical limits to sensation and perception. *Annu Rev Biophys Biophys Chem* 16:455–478.
- Barnes R, Jones MR (2000) Expectancy, attention, and time. *Cogn Psychol* 41:254–311.
- Barrett LF (2009) The future of psychology: connecting mind to brain. *Perspect Psychol Sci* 4:326–339.
- Bergeson TR, Trehub SE (2006) Infants' perception of rhythmic patterns. *Music Percept* 23:345–360.
- Bharucha JJ (1984) Anchoring effects in music: the resolution of dissonance. *Cogn Psychol* 16:485–518.
- Brovelli A, Ding M, Ledberg A, Chen Y, Nakamura R, Bressler SL (2004) Beta oscillations in a large-scale sensorimotor cortical network: directional influences revealed by Granger causality. *Proc Natl Acad Sci USA* 101:9849–9854.
- Burns EM (1999) Intervals, scales, and tuning. In Deutsch D (ed), *The Psychology of Music*. San Diego: Academic Press, pp. 215–264.
- Burns EM, Campbell SL (1994) Frequency and frequency ratio resolution by possessors of relative and absolute pitch: examples of categorical perception? *J Acoust Soc Am* 96:2704–2719.
- Camalet S, Duke T, Julicher F, Prost J (1999) Auditory sensitivity provided by self tuned critical oscillations of hair cells. *Proc Natl Acad Sci USA* 97:3183–3188.
- Cartwright JHE, Gonzalez DL, Piro O (1999a) Nonlinear dynamics of the perceived pitch of complex sounds. *Phys Rev Lett* 82:5389–5392.
- Chen JL, Penhune VB, Zatorre RJ (2008) Listening to musical rhythms recruits motor regions of the brain. *Cereb Cortex* 18:2844–2854.
- Chertoff ME, Hecox KE (1990) Auditory nonlinearities measured with auditory-evoked potentials. *J Acoust Soc Am* 87:1248–1254.

- Choe Y, Magnasco MO, Hudspeth AJ (1998) A model for amplification of hair-bundle motion by cyclical binding of Ca^{2+} to mechano-electrical-transduction channels. *Proc Natl Acad Sci USA* 95:15321–15336.
- Cont A (2008) Modeling musical anticipation: from the time of music to the music of time. Unpublished Ph.D. dissertation, University of Paris 6 and University of California in San Diego.
- Coombes S, Bressloff PC (eds) (2005) *Bursting: The Genesis of Rhythm in the Nervous System*. Singapore: World Scientific Press.
- Crawford JD (1994) Amplitude expansions for instabilities in populations of globally-coupled oscillators. *J Stat Phys* 74:1047–1084.
- Dowling WJ, Harwood DL (1986) *Music Cognition*. San Diego: Academic Press.
- Drake C, Penel A, Bigand E (2000) Tapping in time with mechanically and expressively performed music. *Music Percept* 18:1–24.
- Duke T, Julicher F (2003) Active traveling wave in the cochlea. *Phys Rev Lett* 90:158101.
- Eguiluz VM, Ospeck M, Choe Y, Hudspeth AJ, Magnasco MO (2000) Essential nonlinearities in hearing. *Phys Rev Lett* 84:5232.
- Escabi MA, Schreiner CE (2002) Nonlinear spectrotemporal sound analysis by neurons in the auditory midbrain. *J Neurosci* 22:4114–4131.
- Fitch WT, Rosenfeld AJ (2007) Perception and production of syncopated rhythms. *Music Percept* 25:43–58.
- FitzHugh R (1961) Impulses and physiological states in theoretical models of nerve membrane. *Biophys J* 1:445–466.
- Fujioka T, Large EW, Trainor LJ, Ross B (2009) Time courses of cortical beta and gamma-band activity during listening to metronome sounds in different tempi. *The neurosciences and music III: disorders and plasticity*. *Ann NY Acad Sci* 1169:89–92.
- Gold T (1948) Hearing II. The physical basis of the action of the cochlea. *Proc R Soc Lond B Biol Sci* 135:492.
- Goldstein JL (1973) An optimal processor theory for the central formation of the pitch of complex tones. *J Acoust Soc Am* 54:1496–1516.
- Grahn JA, Brett M (2007) Rhythm and beat perception in motor areas of the brain. *J Cogn Neurosci* 19:893–906.
- Grothe B (2003) New roles for synaptic inhibition in sound localization. *Nat Rev Neurosci* 4:540–550.
- Guckenheimer J, Kuznetsov YA (2007) Bautin bifurcation. *Scholarpedia*, p. 1853.
- Hannon EE, Johnson SP (2005) Infants use meter to categorize rhythms and melodies: implications for musical structure learning. *Cogn Psychol* 50:354–377.
- Helmholtz HLF (1863) *On the Sensations of Tone as a Physiological Basis for the Theory of music*. New York: Dover Publications.
- Hindmarsh JL, Rose RM (1984) A model of neuronal bursting using three coupled first order differential equations. *Proc R Soc Lond B Biol Sci* 221:87–102.
- Hodgkin A, Huxley A (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. *J Physiol (Lond)* 117:500–544.
- Hoppensteadt FC, Izhikevich EM (1996a) Synaptic organizations and dynamical properties of weakly connected neural oscillators I: analysis of a canonical model. *Biol Cybern* 75:117–127.
- Hoppensteadt FC, Izhikevich EM (1996b) Synaptic organizations and dynamical properties of weakly connected neural oscillators II: learning phase information. *Biol Cybern* 75:126–135.
- Hoppensteadt FC, Izhikevich EM (1997) *Weakly Connected Neural Networks*. New York: Springer.
- 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.
- Irino T, Patterson RD (2006) A dynamic compressive gammachirp auditory filterbank. *IEEE Trans Audio Speech Lang Processing* 14:2222–2232.
- Iversen JR, Repp B, Patel AD (2009) Top-down control of rhythm perception modulates early auditory responses. *The neurosciences and music III: disorders and plasticity*. *Ann NY Acad Sci* 1169:58–73.

- Izhikevich EM (2000) Subcritical elliptic bursting of Bautin type. *SIAM J Appl Math* 60:503–535.
- Izhikevich EM (2007) *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*. Cambridge, MA: MIT Press.
- Izhikevich EM, Edelman GM (2008) Large-scale model of mammalian thalamocortical systems. *Proc Natl Acad Sci USA* 105:3593–3598.
- Jackendoff R (2003) *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.
- Jantzen KJ, Oullier O, Marshall L, Steinberg FL, Kelso JAS (2007) A Parametric fMRI Investigation of context effects in sensorimotor timing and coordination. *Neuropsychologia* 45:673–684.
- Jones MR (1976) Time, our lost dimension: toward a new theory of perception, attention, and memory. *Psychol Rev* 83:323–335.
- Jones MR (2008) Musical time. In Hallam S, Cross I, Thaut M (eds), *Oxford Handbook of Music Psychology*. Oxford: Oxford University Press
- Jones MR, McAuley JD (2005) Time judgments in global temporal contexts. *Percept Psychophys* 67:398–417.
- Jones MR, Yee W (1997) Sensitivity to time change: the role of context and skill. *J Exp Psychol Hum Percept Perform* 23:693–709.
- Jones MR, Moynihan H, MacKenzie N, Puente J (2002) Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychol Sci* 13:313–319.
- Joris PX, Schreiner CE, Rees A (2004) Neural processing of amplitude-modulated sounds. *Physiol Rev* 84:541–577.
- Julicher F (2001) Mechanical oscillations at the cellular scale. *C R Acad Sci IV* 2:849–860.
- Kameoka A, Kuriyagawa M (1969) Consonance theory part II: consonance of complex tones and its calculation method. *J Acoust Soc Am* 45:1460–1471.
- Karabanov A, Blom R, Forsman L, Ullén F (2009) The dorsal auditory pathway is involved in performance of both visual and auditory rhythms. *Neuroimage* 44:480–488.
- Kelso JAS (1995) *Dynamic Patterns: The Self-Organization of Brain and Behavior*. Cambridge, MA: MIT Press.
- Kemp DT (1979) Evidence of mechanical nonlinearity and frequency selective wave amplification in the cochlea. *Eur Arch Otorhinolaryngol* 224:370.
- Kern A, Stoop R (2003) Essential role of couplings between hearing nonlinearities. *Phys Rev Lett* 91:128101–128104.
- Kirchner S, Tomasello M (2009) Joint drumming: social context facilitates synchronization in preschool children. *J Exp Child Psychol* 102:299–314.
- Krumhansl CL (1990) *Cognitive Foundations of Musical Pitch*. New York: Oxford University Press.
- Krumhansl CL, Kessler EJ (1982) Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychol Rev* 89:334–368.
- Kuramoto A (1975) Self-entrainment of a population of coupled nonlinear oscillators. In *International Symposium on Mathematical Problems in Theoretical Physics, Lecture Notes in Physics, Vol 39*. New York: Springer, pp. 420–422.
- Kutas M, Hillyard SA (1980) Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207:203–205.
- Langner G (1992) Periodicity coding in the auditory system. *Hear Res* 60:115–142.
- Langner G (2007) Temporal processing of periodic signals in the auditory system: neuronal representation of pitch, timbre, and harmonicity. *Z Audiol* 46:80–21.
- Large EW (2000) On synchronizing movements to music. *Human Movement Science* 19: 527–566.
- Large EW (2008) Resonating to musical rhythm: theory and experiment. In Grondin S (ed), *The Psychology of Time*. Cambridge: Emerald, pp. 189–231.
- Large EW (in press) Dynamics of musical tonality. In Huys R, Jirsa V (eds), *Nonlinear dynamics in human behavior*. New York: Springer.
- Large EW, Crawford JD (2002) Auditory temporal computation: interval selectivity based on post-inhibitory rebound. *J Comput Neurosci* 13:125–142.

- Large EW, Jones MR (1999) The dynamics of attending: how people track time varying events. *Psychol Rev* 106:119–159.
- Large EW, Palmer C (2002) Perceiving temporal regularity in music. *Cogn Sci* 26:1–37.
- Large EW, Snyder JS (2009) Pulse and meter as neural resonance. *The neurosciences and music III: disorders and plasticity*. *Ann NY Acad Sci* 1169:46–57.
- Large EW, Tretakis AE (2005) Tonality and Nonlinear Resonance. *The neurosciences and music II: from perception to performance*. *Ann NY Acad Sci* 1060:53–56.
- Large EW, Fink P, Kelso JAS (2002) Tracking simple and complex sequences. *Psychol Res* 66:3–17.
- Large EW, Almonte F, Velasco M (2010) A canonical model for gradient frequency neural networks. *Physica D: Nonlinear Phenomena* 239:905–911.
- Larson S (2004) Musical forces and melodic expectations: comparing computer models and experimental results. *Music Percept* 21:457–498.
- Lee KM, Skoe E, Kraus N, Ashley R (2009) Selective subcortical enhancement of musical intervals in musicians. *J Neurosci* 29:5832–5840.
- Lerdahl F (2001) *Tonal Pitch Space*. New York: Oxford University Press.
- Lerdahl F, Jackendoff R (1983) *A generative theory of tonal music*. Cambridge: MIT Press.
- Licklider JCR (1956) Auditory frequency analysis. In Cherry C (ed), *Information Theory*. New York: Academic Press, pp. 253–268.
- London JM (2004) *Hearing in Time: Psychological Aspects of Musical Meter*. New York: Oxford University Press.
- MacKay WA, Mendonca AJ (1995) Field potential oscillatory bursts in parietal cortex before and during reach. *Brain Res* 704:167–174.
- McAuley DJ (1995) *Perception of Time Phase: Toward an Adaptive Oscillator Model of Rhythmic Pattern Processing*. Bloomington, IN: Indiana University Press.
- McAuley JD, Kidd GR (1995) Temporally directed attending in the discrimination of tempo: further evidence for an entrainment model. *J Acoust Soc Am* 97:3278.
- Murphy WJ, Tubis A, Talmadge CL, Long GR, Krieg EF (1996) Relaxation dynamics of spontaneous otoacoustic emissions perturbed by external tone. 3. Response to a single tone at multiple suppression levels. *J Acoust Soc Am* 100:3979–3982.
- Murthy VN, Fetz EE (1992) Coherent 25- to 35-Hz oscillations in the sensorimotor cortex of awake behaving monkeys. *Proc Natl Acad Sci USA* 89:5670–5674.
- Nagumo J, Arimoto S, Yoshizawa S (1962) An active pulse transmission line simulating nerve axon. *Proc IRE* 50:2061–2070.
- Ohm GS (1843) Über die Definition des Tones, nebst daran geknüpfter Theorie der Sirene und ähnlicher tonbildender Vorrichtungen. *Ann Phys Chem* 135:513–565.
- Palmer C (1989) Mapping musical thought to musical performance. *J Exp Psychol Hum Percept Perform* 15:331–346.
- Palmer C (1997) Music performance. *Annu Rev Psychol* 48:115–138.
- Pandya PK, Krishnan A (2004) Human frequency-following response correlates of the distortion product at 2F1–F2. *J Am Acad Audiol* 15:184–197.
- Parncutt R (1994) A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Percept* 11:409–464.
- Patel AD (2007) *Music, Language, and the Brain*. Oxford: Oxford University Press.
- Patel AD, Iversen JR, Chen YQ, Repp BH (2005) The influence of metricality and modality on synchronization with a beat. *Exp Brain Res* 163:226–238.
- Patel AD, Iversen JR, Bregman MR, Schulz I (2009) Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Curr Biol* 19:827–830.
- Patterson RD, Robinson K, Holdsworth J, McKeown D, Zhang C, Allerhand M (1992) Complex sounds and auditory images. In Cazals Y, Demany L, Horner K (eds), *Auditory Physiology and Perception*, Proc 9th International Symposium on Hearing. Oxford: Pergamon, pp. 429–446.
- Penel A, Drake C (1998) Sources of timing variations in music performance: a psychological segmentation model. *Psychol Res* 61:12–32.
- Pfurtscheller G, Lopes da Silva FH (1999) Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin Neurophysiol* 110:1842–1857.

- Plack CJ, Oxenham AJ (2005) The psychophysics of pitch. In Plack CJ, Fay RR, Oxenham AJ, Popper AN (eds), *Pitch: Neural Coding and Perception*. New York: Springer, pp. 7–55.
- Plomp R, Levelt WJM (1965) Tonal consonance and critical bandwidth. *J Acoust Soc Am* 38:548–560.
- Poeppel D, Embick D (2005) Defining the relation between linguistics and neuroscience. In Cutler A (ed), *Twenty-First Century Psycholinguistics: Four Cornerstones*. Mahwah, NJ: Lawrence Erlbaum, pp. 103–118.
- Prince A, Smolensky P (1997) Optimality: from neural networks to universal grammar. *Science* 275:1604–1610.
- Provasi J, Bobin-Begue A (2003) Spontaneous motor tempo and rhythmical synchronisation in 2-1/2 and 4-year-old children. *Int J Behav Devel* 27:220–231.
- Purcell DW, Ross B, Picton TW, Pantev C (2007) Cortical responses to the 2f1–f2 combination tone measured indirectly using magnetoencephalography. *J Acoust Soc Am* 122:992–1003.
- Quené H, Port RF (2005) Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica* 62:1–13.
- Rankin SK, Large EW, Fink PW (2009) Fractal tempo fluctuation and pulse prediction. *Music Percept* 26:401–413.
- Repp BH (2002) The embodiment of musical structure: effects of musical context on sensorimotor synchronization with complex timing patterns. In Prinz W, Hommel B (eds), *Common Mechanisms in Perception and Action*. New York: Oxford University Press, pp. 245–265.
- Repp BH (2008) Multiple temporal references in sensorimotor synchronization with metrical auditory sequences. *Psychol Res* 72:79–98.
- Robles L, Ruggero MA, Rich NC (1997) Two-tone distortion on the basilar membrane of the chinchilla cochlea. *J Neurophysiol* 77:2385–2399.
- Rougeul A, Bouyer JJ, Dedet L, Debray O (1979) Fast somato-parietal rhythms during combined focal attention and immobility in baboon and squirrel monkey. *Electroencephalogr Clin Neurophysiol* 46:310–319.
- Ruggero MA (1992) Responses to sound of the basilar membrane of the mammalian cochlea. *Curr Opin Neurobiol* 2:449–456.
- 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.
- Sakai K, Hikosaka O, Miyauchi S, Takino R, Tamada T, Iwata NK, Nielsen M (1999) Neural representation of a rhythm depends on its interval ratio. *J Neurosci* 19:10074–10081.
- Salenius S, Hari R (2003) Synchronous cortical oscillatory activity during motor action. *Curr Opin Neurobiol* 13:678–684.
- Sanes JN, Donoghue JP (1993) Oscillations in local field potentials of the primate motor cortex during voluntary movement. *Proc Natl Acad Sci USA* 90:4470–4474.
- Schachner A, Brady TF, Pepperberg IM, Hauser MD (2009) Spontaneous motor entrainment to music in multiple vocal mimicking species. *Curr Biol* 19:831–836.
- Schellenberg EG, Trehub SE (1994) Frequency ratios and the perception of tone patterns. *Psychon Bull Rev* 1:191–201.
- Schellenberg EG, Trehub SE (1996) Natural musical intervals: evidence from infant listeners. *Psychol Sci* 7:272–277.
- Schouten JF (1938) The Perception of subjective tones. *Proc Kon Akad Wetenschap* 41:1086–1093.
- Schouten JF, Ritsma RJ, Cardozo BL (1962) Pitch of the residue. *J Acoust Soc Am* 34:1418–1424.
- Seebeck A (1841) Beobachtungen über einige Bedingungen der Entstehung von Tönen. *Ann Phys Chem* 53:417–436.
- Seebeck A (1843) Über die Definition des Tones. *Ann Phys Chem* 139:353–368.
- Seth AK, Izhikevich E, Reeke GN, Edelman GM (2006) Theories and measures of consciousness: an extended framework. *Proc Natl Acad Sci USA* 103:10799–10804.
- Shapira Lots I, Stone L (2008) Perception of musical consonance and dissonance: an outcome of neural synchronization. *J R Soc Interface* 5:1429–1434.

- Sloboda JA (1983) The communication of musical metre in piano performance. *Q J Exp Psychol* 35:377–396.
- Sloboda JA (1985) Expressive skill in two pianists – metrical communication in real and simulated performances. *Can J Psychol* 39:273–293.
- Sloboda JA, Juslin PN (2001) Psychological perspectives on music and emotion. In Juslin PN, Sloboda JA (eds), *Music and Emotion: Theory and Research*. New York: Oxford University Press, pp. 71–104.
- Smith JD, Nelson DG, Grohskopf LA, Appleton T (1994) What child is this? What interval was that? Familiar tunes and music perception in novice listeners. *Cognition* 52:23–54.
- Snyder JS, Krumhansl CL (2001) Tapping to ragtime: cues to pulse finding. *Music Percept* 18:455–489.
- Snyder JS, Large EW (2005) Gamma-band activity reflects the metric structure of rhythmic tone sequences. *Cogn Brain Res* 24:117–126.
- Stefanescu R, Jirsa V (2008) A low dimensional description of globally coupled heterogeneous neural networks of excitatory and inhibitory neurons. *PLoS Comp Biol* 4:e1000219.
- Strogatz SH (2000) From Kuramoto to Crawford: exploring the onset of synchronization in populations of coupled oscillators. *Physica D* 143:1–20.
- Sutter ML, Schreiner C (1991) Physiology and topography of neurons with multi-peaked tuning curves in cat primary auditory cortex. *J Neurophysiol* 65:1207–1226.
- Tallon-Baudry C, & Bertrand, O. (1999) Oscillatory gamma activity in humans and its role in object representation. *Trends Cogn Sci* 3:151–162.
- Terhardt E (1974) Pitch, consonance, and harmony. *J Acoust Soc Am* 55:1061–1069.
- Todd NPM (1985) A model of expressive timing in tonal music. *Music Percept* 3:33–59.
- Toiviainen P (1998) An interactive MIDI accompanist. *Comput Music J* 22:63–75.
- Toiviainen P, Snyder JS (2003) Tapping to Bach: resonance-based modeling of pulse. *Music Percept* 21:43–80.
- von Békésy G (1960) *Experiments in Hearing*. New York: McGraw-Hill.
- Vos PG (1973) Waarneming van metrische toonreeksen. *Stichting Studentenpers, Nijmegen*.
- Wiggins S (1990) *Introduction to Applied Nonlinear Dynamical Systems and Chaos*. New York: Springer.
- Wilson HR, Cowan JD (1973) A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik* 13:55–80.
- Winkler I, Haden GP, Ladinig O, Sziller I, Honing H (2009) Newborn infants detect the beat in music. *Proc Natl Acad Sci USA* 106:2468–2471.
- Zuckerkandl V (1956) *Sound and Symbol: Music and the External World*. Princeton, NJ: Princeton University Press.

Chapter 8

Memory for Melodies

Andrea R. Halpern and James C. Bartlett

8.1 Introductory Comments

Memory for music presents a paradox. On the one hand, memory for music that people have already learned can be astonishingly good, both in extent and longevity. On the former point, consider how many tunes an average person could recognize, or even recall. No one has even attempted to measure the limits of musical memory. Concerning longevity, older adults can show excellent retention of music learned decades previously (Bartlett and Snelus 1981; Rubin et al. 1998). Even early-stage Alzheimer's disease patients can almost perfectly discriminate familiar tunes such as patriotic and holiday songs from musically similar but unfamiliar tunes (Bartlett et al. 1995). And this memory can persist not just for songs that have words, but also for purely melodic motives, and without much context. For instance, it is not uncommon to turn on the radio and hear just a few notes of a tune, and be able immediately to hum along or at least recognize the tune as familiar.

Musical memory also shows its persistence by being veridical, or capturing aspects of the music reasonably faithfully. Several researchers have shown that the absolute pitch of familiar music is remembered fairly well, within two semitones, even among nonmusicians and nonpossessors of absolute pitch (Halpern 1989; Levitin 1994; Schellenberg and Trehub 2003), as is tempo (Halpern 1988; Levitin and Cook 1996). Some evidence also suggests that even judgments of musical emotion can be extracted from remembered music similarly to those extracted from sounded music (Lucas et al. 2010). These demonstrations are notable because the identity of music comes from the *relationships* between successive pitches and temporal units, so the memory for absolute tempo and pitch seems to be beyond what is required for making sense of music.

On the other hand, memory for music can be very poor, particularly when learning new music. Typically, memory for music is assessed by recognition, as

A.R. Halpern (✉)

Psychology Department, Bucknell University, Lewisburg, PA 17837, USA
e-mail: ahalpern@bucknell.edu

recall invites difficult issues of production competence. Thus this chapter does not consider the kind of deliberate memorization for later recall required in musical performance. But even the simplest kind of recognition test for melodies shows how poor musical memory can be, in comparison to other kinds of memory. A student was recently setting up a study of recognition memory for paintings. The study session consisted of viewing each of 28 paintings for 3 s, followed by 45 min of visual illusion distraction, and then a surprise old/new recognition test with 28 old and 28 new paintings. Performance was virtually perfect and measures had to be taken to make the task harder. Almost legendary is Standing's (1973) finding that memory for pictures is nearly limitless (10,000 items were presented in that study).

In contrast, Halpern and O'Connor (2000) designed a music recognition memory test that would be feasible for early-stage Alzheimer's patients. Eight novel tunes were presented for incidental encoding, followed immediately by eight old and eight new tunes. Pilot work showed that young adult normal controls could not do this task much above chance levels, necessitating two presentations of the tunes during learning (which brought performance up to a respectable but not overwhelming level). Using longer study sequences, Halpern and Müllensiefen (2008) presented 40 unfamiliar melodies under various encoding conditions, followed by old/new recognition of 80 tunes, using a 6-point confidence scale. Area-under-the-Receiver-Operating-Characteristic (ROC)-curve scores were about 0.70 (0.50 = chance; 1.0 = perfect performance), which is in the range of performance levels in quite a few of the studies reviewed herein. Again, the results are respectable but not spectacular, and far below recognition levels for other rich materials.

This paradox is interesting, most obviously because it raises the question of how new music becomes well learned, if learning is so laborious at first or second exposure. It is also an intriguing puzzle because music is eventually learned even by nonmusicians, who have few analytic strategies to help them, and even for music with few semantic associations or internal references, such as classical themes.

In other domains, variability in learning can be partly accounted for by quality of encoding. In Levels of Processing (LOP) studies (Craik and Lockhart 1972), memory researchers can often increase quality of retrieval by imposing or encouraging elaborative encoding tasks, such as asking people to generate a synonym for a to-be-remembered word. Perhaps music learning is often difficult because listeners do not (or cannot) use elaborative encoding. However, evidence suggests rather that this memory "law" does not seem to obtain in music. As one indication, a recent database search for "Levels of Processing" and "music" turned up virtually no entries. A few studies have shown encoding task effects for well-known tunes (judging familiarity of the tune produced better recognition than judging what instrument was playing the tune; Peretz et al. 1998), but memory for well-known music may also rely on semantic or other nonmusical strategies. Certainly both of the current authors have failed to find LOP effects for unfamiliar music on numerous occasions (some published, some languishing in bottom drawers). Thus it is likely that factors other than conditions of encoding are more important in memory for music than in other domains.

This chapter examines some of the other factors that appear to modulate tune learning. (Note: Many of the studies considered here use simple, single-line melodies, without words. However, a few use fully realized music with orchestration and harmonies, which are pointed out when appropriate). Long-term retention (Sect. 8.2.1) is one focus, as detailed in the preceding text, but another focus is short-term retention such as that needed for immediate same–different comparisons (Sect. 8.2.2). Other factors affecting memory for music include aspects of the tunes, for example, degree of familiarity of the item (Sect. 8.2), as well as familiarity and well-formedness of the musical system from which the tunes are derived (Sect. 8.4). The chapter also considers temporal factors, such as the influence of retention interval on what listeners learn about melodies (Sect. 8.3), as well as two important aspects of listeners themselves: their musical experience (Sect. 8.5) and their age (ranging for current purposes from young to senior adult, Sect. 8.6). It turns out that these last two factors have some expected, but also some unexpected relationships (or absence thereof) with retention of music. The relationship between these two variables is also intriguing, on the supposition that benefits from increased domain-related experience might mitigate some age-related declines in memory. As seen further on, this does not appear to be the case, unfortunately. The chapter concludes with some thoughts on how memory for music may be similar to and different than memory for other kinds of materials.

8.2 Familiarity and Nameability of Melodies

Perhaps the most powerful variable affecting music recognition has been referred to in the literature as familiarity. The term is not ideal, for at least two different reasons. First, in most of the relevant research, familiarity has been operationalized through a comparison of tunes unknown to participants prior to a study with well-known tunes they had heard frequently in life. Although in general the investigators have attempted to avoid confounding “familiarity” with perceptual and musicological features of the stimuli (e.g., tonality, or adherence to a scale, and rhythm), another confounding factor has been less often addressed: that between the extent of prior “real life” exposure to a melody and its verbal identifiability, through, for example, recall of its title, some of its lyrics, or identifying contextual information (“it’s theme song of the musical ‘Cats’”). In the remainder of this chapter, these two aspects of tune knowledge are referred to as “real-life exposure” and “nameability.” One key point that emerges in this discussion is that some of what researchers know about *familiarity* effects might be better characterized as *nameability* effects.

A second problem with “familiarity” is that a wealth of evidence from the human memory literature supports a dual-process theory of memory: the notion that two cognitive processes underlie retrieval, referred to as “familiarity” and “recollection” (see Yonelinas 2002 for a review). Familiarity is viewed as an overall feeling of “oldness” that can vary in strength but lacks any context cues (“I cannot place that tune but it sure sounds familiar”), whereas recollection refers to the conscious

recollection of detailed perceptual and contextual information about a prior experience (“I heard that same song last night at a party”). This state of affairs can lead to mind-bending tongue-twisters (e.g., “familiarity affected both recollection and familiarity”) that can cause confusion. To minimize such confusion, the term “prior knowledge” refers to comparisons of well-known to novel tunes (or musical genres). A distinction is made in cases where a prior knowledge effect might be better characterized as a “nameability effect” as opposed to a “real-life exposure effect.” The terms “recollection” and “familiarity” are used in accordance with the human memory literature, as mentioned previously.

8.2.1 *Long-Term Memory*

A task showing dramatic prior knowledge effects is long-term recognition memory. The most common method of testing such memory is that of presenting a variable-length sequence or list of stimuli, depending on how memorable the stimuli are, followed by a test including “old” items from the study list intermixed with “new” items not heard before. The test typically follows the study phase by 10–30 min, which qualifies this paradigm as testing “long-term” memory, at least in contrast to comparison of two tunes played in succession (see next section). Performance accuracy is typically assessed by examining both hit rates (the proportion of old items called “old”) and false-alarm rates (the proportion of new items called “old”), with a high hit rate and low false-alarm rate signifying good performance. Recognition judgments are substantially more accurate for well-known tunes than for novel tunes (Bartlett et al. 1995). However, some nuances surrounding this basic observation offer valuable clues as to the nature of the processes that support melodic memory.

Bartlett et al. (1995) employed a trained musician to compose a set of novel tunes that matched a set of novel tunes in number of notes, average interval size, rhythmic units and general pleasantness. In two of their experiments, the well-known and novel tunes were presented in separate study lists, each followed by a recognition test. Both young adults and healthy older people (59–80 years old) showed higher hit rates and lower false-alarm rates for the well-known tunes than for the novel tunes, suggesting a difference in recognition accuracy. This pattern is quite often observed in comparisons of easier and more difficult items in recognition memory (Glanzer and Adams 1985), so it was not surprising. What was surprising was the absence of this pattern when the well-known and novel tunes were intermixed in the study lists and tests. In this case the hit rates were dramatically higher for the well-known tunes than for the novel tunes, as was true in the separate lists. However, the false-alarm rates were approximately equal for the two tune types. In terms of signal detection theory, old–new discrimination was much greater for well-known tunes than for novel tunes, but there also was a bias to judge the well-known tunes as “old.”

What might it mean that the intermixed list of novel and well-known tunes prevented people from suppressing false alarms to the well-known tunes? One plausible

hypothesis is that old–new judgments in tune recognition are based to a substantial extent on subjective familiarity, in the absence of recollection of information specifying the source of the familiarity (e.g., the studied items versus last year’s Christmas party). Familiarity will be much stronger for well-known tunes than novel tunes, and this will tend to increase the hit rate advantage of well-known tunes, while possibly increasing false-alarm rates for those same well-known tunes. In a between-list design, where well-known and novel tunes are presented and then tested separately, listeners can easily compensate for this tendency by adopting a more stringent recognition criterion for well-known tunes than novel tunes. In other words, listeners might only say “old” to a well-known tune if the tune seems *very* familiar. In a within-list (intermixed) design, however, this would be harder to do as the listener would need to adjust that criterion trial by trial. There is substantial evidence that participants often fail to adjust their recognition criteria for individual items in a single recognition test (see Benjamin 2008 for a review).

Some findings of McAuley et al. (2004) underscore the importance of familiarity in the absence of recollection, in recognition memory for tunes. These investigators compared memory for novel and well-known melodies in a variant of the standard recognition task designed to test knowledge of how recently and how frequently tunes had been studied. The novel melodies were composed for the experiment in a range of major and minor keys, rhythms, speeds, and melodic contours, with the goal that they would be at least as distinctive as the well-known melodies and approximately as long (mean = 12.3 notes versus 15.6 notes for the well-known tunes). The participants heard a sequence of novel and well-known melodies in which half of the items were presented one time and the others were presented three times. One day later, they heard a second sequence of (different) melodies, constructed in the same way. The second list was followed by two memory tests, one in which subjects judged the frequency of items (one versus three presentations), and a second in which they judged the recency of items (same day versus previous day). Frequency judgments were slightly less accurate for novel tunes than for well-known tunes, but discrimination between thrice-presented items and once-presented items was well above chance for both. By contrast, recency judgments were substantially less accurate for the novel tunes than for well-known tunes, and discrimination between day-1 items and day-2 items approximated chance for the novel tunes. Moreover, the recency judgments to novel tunes were affected more by frequency than by recency itself. That is, thrice-presented tunes heard on day 1 received more “same day” judgments than did once-presented tunes heard on day 2. These findings suggest that, in the case of novel tunes, time of presentation is poorly recollected and that memory judgments are based for the most part on familiarity strength.

What about judgments to well-known tunes? The improved recency judgments with well-known tunes suggest that recollection is greater with such tunes than with novel tunes. However, McAuley et al. performed an analysis suggesting that this difference reflects the often high nameability of well-known tunes rather than the fact that they have been experienced in life. Specifically, the authors found a reliable positive correlation between the accuracy of recency judgments to well-known

tunes and the nameability of these tunes (based on a naming test administered to each participant at the end of the experimental session, $r = 0.52$). Hence, the recollection advantage of well-known tunes is due not simply to the fact that they are known; it depends on nameability.

A link between the nameability of tunes and the process of recollection has also been supported in a study that actually tested melodic recall, unlike experiments considered heretofore that tested simply recognition and related judgments (frequency and recency) Using a unique methodology, Korenman and Peynircioğlu (2004) presented tunes paired with animal names, followed by tests of (1) recall of the animal names in response to the melodies and (2) recall of the melodies in response to the animal names (the hummed responses were recorded and later scored). Participants in three different groups received: (1) original recordings of well-known melodies with full orchestration, (2) single-line versions of these same melodies played on a synthesizer, and (3) single-line versions of unknown melodies played on the synthesizer. The same animal names, paired at random with the melodies, were used in all conditions.

The recall results were straightforward: Recall of names in response to melodies was better than recall of melodies in response to names, perhaps because it was easier to guess a correct name than to guess (through humming) a correct melody. Correct recall was substantially higher for well-known tunes than for novel tunes, despite the fact that the study list was shorter in the novel-tune condition to minimize floor effects. As in the McAuley et al. (2004) study, the authors assessed knowledge of the names of the well-known tunes in the last phase of the study. Associative recall (both melody-from-name and name-from-melody) in the group with above-average knowledge of the tune names (“experts”) was approximately twice that in the group less knowledgeable about the tune names (“nonexperts”), indicating that nameability of melodies is an important factor in recollecting contextual information. In considering these data, it is important to remember that the participants were tested on their memory for new associations between melodic snippets and animal names, not actual tune titles or lyrics. Thus, the findings indicate that recollecting the verbal context of a melody’s presentation – or the melodic context of a word’s presentation – is better if the melodies are well known, and especially if they are nameable.

It is interesting to note that when the participants in the Korenman and Peynircioğlu (2004) study could not recall an animal name in response to a melody or a melody in response to a name, they estimated their chances of recognizing the association (i.e., they made a “feeling of knowing” judgment). Moreover, all of the participants were subsequently tested on associative recognition (i.e., they attempted to select which of three names belonged with each of a set of melodies and which of three melodies belonged with each of a set of names). Neither feeling-of-knowing ratings nor associative recognition differed between well-known and novel tunes or between “experts” and “nonexperts” with the well-known tunes. This finding is important because it demonstrates that the nameability of a tune does not affect memory for contextual information so long as the pairing of a tune and its context at study are reinstated at test (as in an associative recognition test). Rather,

the effect of nameability is on *recollection* of contextual information not physically available at test.

In sum, the evidence suggests that recognition of well-known tunes differs from recognition of novel tunes in two important ways. First, well-known tunes create a stronger feeling of familiarity, and because familiarity is an important basis for old–new judgments in recognition memory, participants show a bias to judge well-known tunes as “old” (i.e., heard previously at study). They show this bias only in within-list designs presumably because, in between-list designs, they are able to use a more stringent criterion for recognizing well-known tunes than for recognizing novel tunes. When such criterion adjustments are difficult, as they are in within-list designs, the high familiarity of well-known tunes often leads to “old” judgments even when these tunes are new. Second, many well-known tunes are more nameable than novel tunes, and nameability is linked to the power of tunes to spur recollection of contextual and associative information. Recollection is a hallmark of “episodic memory” (Tuvling 1983), a major component of memory – possibly involving a dedicated brain system (Schacter and Tulving 1994) – that mediates our ability to consciously re-experience events from our personal pasts. Performance in tests of episodic memory are seriously impaired in amnesic patients who have suffered damage in medial–temporal and prefrontal brain regions, and McAuley et al. (2004) make the interesting observation that the memory performance of healthy adults with unfamiliar (and unnameable) tunes resembles that of amnesic individuals with well-known words (i.e., frequency and recency are confused). The brain processes of episodic memory, presumably intact in healthy adults tested by McAuley et al., cannot be engaged in the processing of tunes that cannot be uniquely identified or named.

An alternative hypothesis holds that hard-to-identify tunes engage episodic memory processes, but suffer with respect to elaborative encoding. A wealth of evidence suggests that successful recollection in tests of episodic memory depends on elaborative encoding when the item is first presented (Yonelinas 2002, and see Sect. 8.1), and that such elaborative encoding aids in the creation of distinctive representations that yield good recollection because they are less confusable with other memories at retrieval (see, e.g., Eysenck 1979). Nameability may improve elaborative encoding of the type that produces distinctive and retrievable memory codes. For example, an elaborative encoding of an unknown tune might include information that it sounds very pleasant and might make a good Christmas carol. However, such an encoding is likely to be applicable to several different tunes in a study sequence, and so it is not distinctive. By contrast, elaborative encoding of a nameable tune might be highly distinctive (e.g., “that was mother’s favorite Christmas carol”), supporting recollection in a subsequent test.

How familiarity and nameability are linked to episodic memory is an important issue for future research to address. However, it likewise is important to understand the processes underlying the detection that a tune is familiar and the retrieval of its name. Dalla Bella et al. (2003) explored this issue by presenting the beginnings of known and novel tunes in a “gating” paradigm in which listeners first heard the first note of a tune, then the first two notes, then the first three notes, and so on until they

judged the tune to be familiar with high confidence on each of three successive trials. The known melodies had been previously classified as “highly familiar” or “moderately familiar” in a prior norming study. High-confidence identification of tunes as familiar occurred after six notes (on average) for the “highly familiar” tunes and about eight notes for the “moderately familiar” tunes. Similar results were obtained in a second experiment in which tune identifications were based on singing continuations with accuracy and high confidence on three successive trials. These analyses were based only on tunes that were, eventually, successfully recognized (Experiment 1) or sung (Experiment 2), and so the findings suggest that even when tunes are known by a listener, they can be identified more quickly if they are more “familiar.” In light of the preceding discussion, it is important to learn if this “familiarity” effect is one of nameability or merely real-life exposure. It also is important to know whether a tune that sounds familiar and yet cannot be named can nonetheless be uniquely identified through singing its melody. In the Billy Joel song, “The Piano Man,” the denizens of a bar sing out an old song for which they cannot recall either title or lyrics (due perhaps to their level of intoxication), raising the hypothesis that nonverbal identification (through singing) and verbal identification (through naming) might be dissociable.

8.2.2 *Short-Term Memory*

Another popular paradigm for studying melodic processing is the short-term same–different task in which two short melodies are presented in succession, with the second (comparison) matching or mismatching the first (standard) in some designated way. If the melodies are short – say, five to seven notes long – and the task is simply to judge whether the two melodies are physically identical, as opposed to having one or two notes changed, performance will be near the ceiling. However, the task is more difficult if the melodies are longer, or if they are presented at extremely fast or slow tempos. Another difficulty ensues if the two melodies of a pair begin on different notes and the task is to judge whether, despite the change in absolute pitch levels, the second is an accurate transposition of the first into a different key. This transposition detection task requires the processing of *pitch interval* information, as opposed to *absolute pitch* information, as only the former remains constant when a tune is transposed.

When the same–different task is made difficult in any one of the aforementioned ways, prior knowledge of melodies has very large effects. In one recent study, Dowling et al. (2008) asked their listeners to compare well-known and novel melodies 11–21 notes in length in a short-term same–different task, presenting the melodies at extremely fast, medium, or extremely slow tempos (0.6, 3.0, and 6.0 notes/s, respectively). To make the task even more challenging, the different trials involved changes in only two notes. The largest effect in the study was that of prior knowledge, with area-under-ROC scores averaging 0.85 and 0.63 for the

well-known and novel tunes, respectively. The knowledge by tempo interaction was reliable as well, reflecting the fact that, although an advantage for well-known tunes was everywhere apparent, it was stronger at the medium tempo (which was approximately the familiar tempo for the tune) than at the fast and slow tempos. This was a surprising result, as the intuitive prediction was that fast or slow presentation would produce the greatest difficulty with unfamiliar tunes. However, the result should be viewed in the context of prior evidence that identification of well-known tunes is impaired at fast and slow tempos (Warren et al. 1991; Andrews et al. 1998). Thus, if the advantage of well-known tunes results from the fact that they are nameable, it makes sense that fast and slow presentations, which reduce nameability, should reduce the advantage.

Strong effects of prior knowledge in short-term memory have also been found in transposition detection (see Dowling 1982 and Dowling and Harwood 1986 for reviews). This task is a good test of pitch-interval processing, as that information is the same no matter what the starting note is (“Happy Birthday” is the same melody with the same pitch intervals regardless of what pitch someone begins with). If the standard and comparison melodies both are novel and also share melodic contour (the sequence of ups and downs in pitch), the task is quite hard, even for persons with musical training (though more musical participants do perform somewhat better). Indeed, if the standard and comparison are in the same or closely related keys, discrimination of exact from inexact transpositions is close to chance (Bartlett and Dowling 1980). With well-known melodies, however, the task is quite trivial, with even musically untrained participants performing near ceiling. For example, if the standard melody is from a well-known tune (e.g., the first phrase of “She’ll Be Coming Around the Mountain”), and the comparison is a transposed version with one note changed, participants almost always detect the difference: the comparison is perceived as simply not the same song. This is notable in light of work suggesting that monkeys will accept a transposed tune as the same as an original only if the notes have been changed by exactly an octave (Wright et al. 2000).

What does the effect of prior knowledge on transposition detection tell us about melodic processing? Since the transposition detection task poses a minimal load on short-term memory (again, as long as the melodies are short) the clearest implication concerns the process of perceptually encoding the precise melodic intervals that – along with rhythm, meter (for instance a 2-beat march versus a 3-beat waltz) and a few other factors – distinguish one song from another in our culture. Such encoding is apparently quite difficult the first few times a novel melody is heard, and yet it is eventually accomplished for all the tunes that people know well. Deutsch (1979) has shown that transposition detection with novel tunes improves if the first tune in each pair is presented six times as opposed to only once. Beyond this, however, almost nothing is known about the time course of pitch interval encoding as a tune progresses from being completely novel to being very well known.

In fact, researchers do not even know how to characterize the codes that capture pitch-interval information at different levels of learning and musical expertise. In some cases, interval information might consist of something akin to the ratios of frequencies between successive notes (inter-note interval information). In other cases, however, interval information might be encoded in terms of steps on the diatonic scale (this is the *do-re-mi* scale that many learn in childhood). Diatonic scale-step information is referred to as chroma information and is contrasted with “pitch height” information in the literature. Thus, when the note C is played in different octaves, pitch height changes but chroma remains constant (Shepard and Jordan 1984; Dowling and Harwood 1986; Dowling et al. 1995). Chroma encoding is used in recognition of well-known tunes (see Dowling 1991 for a review). For example, well-known tunes can be recognized when the pitches of individual notes have been manipulated by transposing them up or down by one octave, maintaining chroma while drastically altering pitch height (Idson and Massaro 1978). However, good recognition of such octave-manipulated melodies depends on their maintaining correct melodic contour, which means that the code used to recognize such melodies is more than simply a sequence of chromas. Specifically, the code must contain some information about inter-note pitch intervals, though this information might be global and not very precise (e.g., it might be melodic contour, the sequence of rises and falls in pitch height). Along similar lines, the transposition detection study by Deutsch (1979) included a condition in which the successive notes of the standard melody were placed in different octaves across six presentations. Surprisingly, performance in this octave-scrambled condition was actually worse than in the single-presentation (and unscrambled) condition, and substantially worse than in the unscrambled six-presentation condition. This finding suggests it is difficult to learn the pitch-interval structure of novel melodies through the encoding of chromas alone. However, encoding of melodies based on chroma and contour appears to provide a viable account of the data in hand (Dowling 1991).

In summary, people seem to remember some aspects of melodies reasonably well over the short term. As mentioned previously, simple same–different judgments to pairs of novel melodies of five to seven notes are made with high accuracy when transposition detection is not required, suggesting highly accurate short-term memory for several different pitches. Second, whereas transposition detection with novel melodies is highly error-prone, it can be greatly improved if the “different” trials involve changes in contour (e.g., if the third interval is rising in the first melody and falling in the next), suggesting that a general up and down pitch pattern is encoded easily (at least if the pattern is relatively simple; see Boltz et al. 1985). Finally, participants appear to be highly sensitive to whether changed notes in the second melody violate the key of the first (Dowling 1978; Bartlett and Dowling 1980), again suggesting that a general sense of scale is encoded fairly well after a short exposure. Thus, it is not that novel melodies are generally hard to encode. Rather, it is the precise pitch interval information in novel melodies that is a source of difficulty. How this difficulty can be overcome – as certainly it is when a tune is well learned – is an important unknown.

8.3 Short-Term Versus Long-Term Memory

After reading Sect. 8.2, the reader may be struck by the very different nature of the questions and methods involved in studies of knowledge effects in short-term memory versus long-term memory. Indeed, little attention has been paid to the short-term-memory/long-term-memory distinction by music cognition researchers. This is unfortunate, as there are indications that the information retained about melodies might be quite different in the two kinds of tasks. The role of contour information, in particular, appears to be different, as suggested by studies by and DeWitt and Crowder (1986) and Dowling and Bartlett (1981). These investigations showed that whereas melodic contour is a salient property of tunes in conditions of immediate testing, even brief filled intervals between a standard melody and a comparison melody greatly reduce its importance. Specifically, if a musically filled interval of even just a few seconds separates a standard tune from a same-contour comparison, listeners have difficulty detecting that their contours match. In line with this observation, the Idson and Massaro (1978) study using scrambled melodies found poor identification of well-known tunes if their chromas were altered, even if their contours were retained. Hence, while contour information can contribute to tune recognition when note chromas are preserved (a point made earlier), contour by itself is a weak cue for recognition in long-term memory tasks.

Although melodic contour appears less important in long-term memory than in short-term memory, this conclusion may depend on defining contour narrowly in the traditional way, as the sequence of ups and downs in pitch within a melodic phrase. Jones et al. (1987) have argued for a broader view of contour which they term “dynamic shape.” Dynamic shape includes rhythmic information as well as melodic ups and downs, and reflects those points in a melody that are attentionally more salient. In support of this view, Jones et al. showed that if a set of study tunes differ in rhythm, lures that match targets in both contour and rhythm attract substantial numbers of false alarm errors. Further, they obtained this result in a long-term memory task across three different levels of initial learning. A subsequent study expanded this result to more familiar melodies (Jones and Ralston 1991).

Whereas melodic contour (as traditionally defined) appears less important in long-term memory than short-term memory, the reverse may be true for interval information. Using a variant of the short-term same–different task, Dowling et al. (2002; see also Dowling et al. 1995) found that discrimination between target melodies and same-contour lures actually improved over a musically filled interval of 5–15 s. By contrast, discrimination between targets and different-contour lures remained roughly constant. This result may suggest that pitch-interval information needs time for consolidation in memory (see, e.g., Patel 2008). Another possibility is that listeners use different codes for interval information in short-term memory versus long-term memory. Note that contour information can be extracted from a sequence of actual inter-note intervals that maintain exact pitches, but not from a sequence of chromas. Hence, if listeners use inter-note interval codes to maintain melodic information in short-term memory tasks, this could explain why they are highly sensitive to contour in these tasks.

Apart from these interpretive issues, an important implication of the Dowling et al. (2002) study is that the classic short-term same-different task requiring transposition detection may underestimate the encoding of interval information into long-term memory. Testing after filled delays may be required to assess the extent of such encoding. Of course, contour and interval are only two types of information that might change in importance, function, and/or representational format between short-term melodic memory and long-term melodic memory. Hébert and Peretz (1997) compared recognition of well-known tunes when pitch interval information had been removed by playing all notes at the same pitch, and when rhythmic information had been changed by playing all notes for the same duration. Performance was much better in the latter condition, suggesting that interval information is more important than rhythm for tune recognition in long-term memory. Given the difficulty that listeners have in the initial encoding of pitch interval information (as revealed in transposition detection tasks), it is not at all clear that the analogous conclusion would hold in immediate short-term memory. Rhythm is also maintained over the long term to some extent, as shown by the fact that performance was best when both types of information were available in the Hébert and Peretz (1997) study. Schulkind (1999) showed that many rhythmic manipulations diminished long-term recognition performance. In light of recent evidence that long-term memory representations contain information about “absolute” musical properties such as pitch and tempo (Halpern 1988, 1989; Levitin 1994; Levitin and Cook 1996; Schellenberg and Trehub, 2003), it is important to compare the roles of such properties in short-term and long-term memory tasks.

8.4 Tune Structure

People remember better items that make sense to them. Tunes can “make sense” (or not) in two major ways. The first way is adherence to tonality. Most music that most people listen to is tonal: Notes and implied or realized harmony conform to a diatonic (musically logical) scale structure. In other words, in most melodies, most notes stay inside in the key of the piece. Sometimes composers violate tonality for aesthetic reasons, such as was true in the 12-tone movement, but not many listeners find that genre appealing. Listeners seem to prefer melodies the more closely they conform to a tonal structure (Cross et al. 1983). This scale structure also facilitates musical processing, as notes are not processed one by one, but as part of a hierarchy of tonal relationships (Dowling 1978).

The second way that tunes can make sense to listeners is if they conform not to just any tonal system, but to *the listeners’* tonal system. In other words, cultural familiarity with a tonal system may facilitate initial processing and thus retention. This second point is different from the first because atonal materials conform to no system, implying that continued exposure to atonal music would not significantly improve processing to such sequences. In contrast, cultural familiarity is assumed

to be an entirely environmental effect, as seen by cross-cultural and some developmental evidence.

The effects of tonality on melody recognition have been studied by several researchers. The typical format for studies varying tonality is short-term transposition detection, one of the tasks described earlier. A common finding is that tonal sequences yield more successful retention over a brief period. For instance, Cuddy and Lyons (1981) presented a standard melody followed by a correct and an incorrect transposition in which one note (and thus two intervals) were changed. Listeners were best able to distinguish these for highly tonal melodies, and were less adept for sequences with ambiguous tonalities. Halpern et al. (1995) compared tonal to atonal sequences in a similar paradigm, although only one comparison was presented at a time. Recognition performance was higher for tonal than for atonal sequences, regardless of whether the tonality manipulation was between or within subjects. Using slightly longer delays in a continuous running memory paradigm (for every melody, say whether it is old or new; some melodies are repeated in the list), Dowling et al. (1995) also found tonal sequences were superior to atonal, but only in the more challenging versions of the task where the delays were filled with other melodies.

On closer inspection, it turns out that the beneficial effects of tonal melodies are not uniform across variations in the task. In the Cuddy and Lyons (1981) study, the changed note in the different comparison sequence did not change the contour of the melody; in the other two studies, new notes (and thus intervals) that changed the contour were compared to note changes that did not change the contour. These latter two studies showed that tonality and type of discrimination interacted: tonality made a difference only when contour was preserved so that the sizes of intervals (as opposed to the directions of intervals) needed to be detected. Performance on changed-contour sequences was not sensitive to tonality. This pattern suggests that the processing benefit of well-formed melodies remembered over short time intervals may be particularly marked when listeners are discriminating fine pitch interval changes rather than coarse contour features of melodies.

Another interesting commonality between the Cuddy and Lyons (1981) study and that by Halpern et al. (1995) is that both tested participants with varying levels of musical training. No interactions of tonality and training were observed. This suggests that nonmusicians have abstracted the orderliness of the tonal system, and use it to increase processing fluency in these discrimination tasks. Recent evidence suggests that some aspects of tonality are processed preattentively even by nonmusicians. Brattico et al. (2006) found that nonmusicians show a robust early negative Event-Related Potential (ERP) response to tunes containing an out-of-key note, even when they were not paying attention to the tunes.

It would be useful to know if tonality confers benefits in retention of melodies over longer time intervals than are typically tested in the laboratory, given that the music that most people eventually learn and retain is highly tonal. No doubt such a task would be aversive to listeners, and perhaps many would predict that tonal items would yield superior memory. But the finding is hardly a foregone conclusion, as false-alarm rates might be higher for new melodies that are tonal versus those that

are atonal, a result that would suggest that if a melody matches well with diatonic scale structure, it feels more familiar (viz. Sect. 8.2). In addition, tonality effects might differ depending on whether tonality is varied between subjects or within subjects. Exposure to a pure list of atonal or weakly tonal items might encourage list-specific strategies, for instance, a note-by-note encoding strategy, because higher-order strategies such as chroma encoding would be ineffective with atonal melodies.

The other kind of musical structure is familiarity with a musical system, defined either as broadly cultural (Chinese versus Western scales) or as a specific idiom (classical or jazz). It seems reasonable that people would use the schemata of their “native” musical tongue to facilitate memory, but few studies have looked at this. Gardiner and Radomski (1999) presented Polish and English listeners with a list of single-line melodies from familiar folk songs from each culture. In immediate recognition, Polish and English listeners were better at discriminating old from new tunes in their own versus the other culture, but only for old responses that were definitely “remembered” (listeners had a clear recollective experience) versus “known” (listeners could say only that they knew the item to be old, but without any clear memory of having heard it). In terms of dual-process theories of human recognition memory, the finding may indicate that if melodies fit well with the musical idiom that a listener has internalized, this improves those processes underlying recollection, but not those that support familiarity.

The familiar music in Gardiner and Radomski’s (1999) study was familiar both culturally and also because the melodies were well known. Demorest et al. (2008) tried to isolate cultural familiarity by using fully realized but unfamiliar music in a cross-cultural study. They recruited listeners in the United States and Turkey. Both groups were presented short lists of excerpts of unfamiliar classical music from Western and Turkish musical traditions, followed by a recognition test. The styles were blocked, and foils were carefully matched to targets in musical aspects. Another test used classical Chinese music. The three musical cultures use different scale systems. US and Turkish listeners were presumed to be unfamiliar with Chinese musical systems, although the Turkish listeners were somewhat familiar with Western music. The authors found a crossover interaction whereby listeners remembered excerpts from their own culture (US or Turkish) better than those from the other culture. Chinese music was recognized poorly by both groups. Turkish listeners did perform better on Western compared to Chinese melodies, consistent with their exposure to Western music (US listeners were equally poor on the nonnative tunes). Musical training did not moderate any of these effects.

Lynch and Eilers (1992) also varied the familiarity of the musical context to look at its effect on detection of mistunings. Although a perception rather than a memory test, they confirmed that adult nonmusicians could detect mistunings quite well in a familiar major scale context, and performed equally poorly on melodies using a novel scale pattern based on augmented intervals and on melodies using an unfamiliar Javanese scale. Interestingly, 1-year-olds showed a pattern similar to that of adults, whereas 6-month-olds performed equally on the major and augmented melodies (and were worse on the Javanese). The authors suggest that musical

acculturation can proceed quickly between 1 and 12 months, but it is clear that some acculturation is in place by 6 months given the poor performance for the Japanese melodies in all age groups.

In summary, it seems that exposure to a body of music that conforms to a particular scale system or style engenders schematic knowledge of the underlying structure of the music. This knowledge can be used to assist encoding of tonal and culturally familiar music, yielding a memory superiority. It is remarkable that only incidental exposure is necessary for these effects to emerge, as the familiarity of the musical system does not seem to interact with musical experience; indeed the Lynch and Eilers (1992) study showed that 1-year-old infants show this schematic knowledge. Nearly universally, musical exposure is widespread, from infant-directed singing to communal activities such as religious services and school assemblies, to the nearly ubiquitous use of electronic musical playback devices among young people in contemporary developed societies. This last point leads to a consideration of what additional benefits in remembering music are associated with deliberate musical training.

8.5 Musical Experience

It is a common, and not unfounded, belief that experts should remember material in their domain better than nonexperts. Indeed, some classic studies have shown that as long as the material is well structured, experts exceed nonexperts in domain-specific memory in such varied domains as chess (Chase and Simon 1973) and figure skating (Deakin and Allard 1991). It turns out that although musical experts exceed nonexperts in some aspects of remembering music, frequently this outcome does not occur.

First, a methodological note: Different studies define musical expertise differently. In some countries, national music competency exams allow a uniform classification scheme. However, other countries such as the United States do not have national exams. Frequently, researchers use years of musical experience (often further defined as music lessons) as the metric for musicianship. This is typically instantiated in forming a group of musicians and one of nonmusicians, but sometimes years of training is used as a covariate. Rarely do researchers actually give musical competency tests before an experiment, allowing years of music lessons (experience) to serve as a proxy for accomplishment (expertise). In some situations, performing experience is counted in lieu of lessons, for instance for jazz musicians, some of whom are largely self taught. Finally, it is standard practice to exclude possessors of absolute pitch, unless that is the topic of interest.

Surprisingly few studies have examined old–new recognition as a function of experience, defined in any way. Two studies mentioned earlier are relevant here. McAuley et al. (2004) presented familiar tunes one or three times, on two successive days, and then asked musicians and nonmusicians for frequency and recency judgments. They found that musicians did not outperform nonmusicians in any

condition. Korenman and Peynircioğlu (2004) failed to find experience effects on either memory or metamemory judgments for musical recognition. As one recent exception to the general findings, Mungan et al. (submitted) presented 24 familiar tunes to trained and untrained listeners, followed by old–new recognition. This was a particularly large sample of 48 people per group, and thus may have been particularly sensitive, but the musicians were superior to nonmusicians in this task. Their advantage occurred not in the hit rate, but in a lower false alarm rate than that of the nonmusicians.

A few studies have embedded this basic task within a more complicated design. For instance, Halpern et al. (1995, Experiment 2) presented 24 unfamiliar melodies, each four times in three different keys, for ratings on pleasantness. Thereafter, old and new items were presented in a short-term same–different task (described later), but participants were asked at that point to indicate old–new recognition for each item as well. Musicians and nonmusicians were both young adult and senior citizens. No effects of musical experience on recognition memory emerged, once vocabulary score was entered as a covariate. In a similar vein, Halpern and Müllensiefen (2008, Experiment 2) presented 40 unfamiliar melodies for later recognition from among 40 new items. This sample had a range of musical experience background, but no effect of years of training as a covariate emerged. In a study previously mentioned, Demorest et al. (2008) found no differences between musicians and nonmusicians on recognition memory of culturally familiar versus unfamiliar songs.

Experience differences are more commonly tested, and found, in short-term musical recognition judgments. As one example, Mikumo (1992) presented tonal or atonal standards to listeners, followed by 12 s of various interference conditions, and a target that could differ from the standard by being an exact transposition, a change of one note in the comparison (but preserving contour), a change of two notes to violate contour, or the comparison was a completely different melody. Musicians outperformed nonmusicians overall, but particularly in the transposition condition, where nonmusicians made many false alarms. Radvansky et al. (1995) presented tonal but unfamiliar tunes as standards, followed by 30 s of a working memory task, then a target that was melodically similar or not to the standard. Half the items also changed timbre. Musicians outperformed nonmusicians in identifying the melodically similar target (timbre change did not affect either group).

In a somewhat more elaborate version of short-term recognition, Halpern et al. (1995) presented standards that were transposed to three keys (for a total of four presentations), followed by a 6-s silent interval, and then a target that was yet another exact transposition, or that changed two of the seven notes. Sometimes the two new notes changed the contour and sometimes they did not. In addition, sequences could be tonal or atonal. The task was to discriminate exact transpositions (same) from inexact (different) ones. In several experiments, musicians were superior to nonmusicians, but only in the condition wherein contour was left unchanged so that changes in exact intervals needed to be monitored. Musical experience was not an advantage when a change of contour was the cue to a different

trial. Musicians were not differentially superior to nonmusicians on tonal or atonal materials.

Another example of short-term recognition was seen in the previously described study by Dowling et al. (2008), which presented pairs of familiar or unfamiliar tunes for comparison at very slow, medium (normal), or very fast tempos. Some trials were exact repetitions and different trials changed two notes with a preserved contour. Musicians were superior to nonmusicians in all conditions in discriminating exact from changed repetitions, including in the easiest condition of comparing two familiar songs at normal speed. This result concurs with the previously mentioned study of tune identification by Dalla Bella et al. (2003), who found that musicians identified well-known and moderately familiar melodies in fewer notes than did nonmusicians, suggesting that training might increase the efficiency of tune identification as a general rule, and not specific to challenging conditions.

To sum up, these studies all suggest that the primary advantage of musical training in remembering melodies occurs when the task requires participants to make fine musically relevant distinctions such as those of interval size, as might occur during a piece when a composer presents variations of a theme. Nonmusicians are quite capable of detecting contour change, which one could argue does not involve such fine musical discrimination as transposition detection. Skills in making fine musically relevant distinctions among melodies are typically tested over short retention intervals and thus within a span of working or short-term memory. Hence, it is unknown whether musical training would confer an advantage in making these same distinctions in long-term episodic memory tasks. The literature seems to show that nonmusicians are as capable as musicians in tests of long-term episodic memory for tunes, but these tasks have typically used simplified materials and have not required the types of subtle musically relevant discriminations required in the short-term memory studies where tunes are presented in quick succession. That musicians appear to be better at identification of tunes known from life requires more research attention, but it may suggest that certain fine discriminations (of precise musical intervals, for example) can facilitate discrimination of such tunes from unknown tunes.

8.6 Aging

One of the issues that has interested the current authors for some time is how music cognition, including memory, changes in normal and pathological aging. This interest stems from both everyday and theoretical bases. It is evident that many older people enjoy music of many genres, as listeners, performers, and financial patrons. In fact, performing arts personnel refer to a “Q-tip Effect” at concerts of jazz, Big Band, or classical music, describing the view from the stage when the spotlights shine through the gray-haired audience. Community bands and orchestras often count senior citizens among their most avid participants, and many retirement communities and nursing homes offer musical activities as part of enrichment and therapy.

Yet very little research has been conducted on this topic. This dearth is regrettable because of some interesting if not unique perspectives that using music as a domain can bring to the study of cognitive aging. For instance, music without words is completely nonverbal yet conveys messages such as valence and arousal (Lucas et al. 2010), making a useful contrast with the large majority of studies in cognitive aging that use language to convey messages. Except for musicians practicing for a concert, most music is learned incidentally, allowing researchers to examine how both particular pieces of music, and the underlying musical structures, may be learned by mere exposure over the lifetime. In addition, musical training can vary at any age, making it possible to separate effects of years of exposure and deliberate training. That is hard to do in most other domains. Finally, one can examine whether music memory is more preserved in pathological aging, such as Alzheimer's disease, compared to well-known verbal impairments.

In most studies of cognitive aging, older adults are defined as 60+ years. In some studies, age is grouped into two or three levels. At other times, age can be used as a continuous variable for somewhat more statistical power. When possible, researchers administer at least one cognitive test not related to music, such as a vocabulary test, to help ensure that any age-related impairments are not attributable to general cognitive decline. This section mostly concerns explicit memory for newly learned material in normally aging adults, where the rememberer is aware that he or she is engaging in an attempt to remember the material, but touches on some other forms of memory and pathological aging as well.

The first point to consider is semantic memory for music, or general memories about music not tied to a specific learning experience. Do older adults remember music learned as younger adults? Several studies agree that familiar music, once firmly encoded, seems retrievable decades later. As noted earlier, Bartlett et al. (1995) presented familiar and novel songs in a recognition task to young adults, normal older adults, and Alzheimer's disease (AD) patients. As part of verifying the stimuli, all groups were asked at the end of the experiment to classify the tunes as familiar or unfamiliar, and to name each one or at least give a descriptor or a few lyrics from the song. The familiar songs were selected to be "lowest common denominator" songs that most Americans would likely learn in childhood, such as patriotic and folk songs. The young and older adults were nearly perfect in classifying the songs, and also scored highly in naming or describing these songs. In fact, this kind of memory seems very robust, as the AD group was very adept in the classification task (they had more problems naming the songs, which is consistent with naming deficits in AD). All three groups were perfect in calling well-known songs "familiar," and they showed a low false alarm rate (occasionally a novel tune was called familiar; the novel tunes were in fact permutations of the familiar tunes, so the occasional false alarm should not be surprising.)

A few other studies have shown that memory for popular music that is first learned in youth seems particularly robust to aging. Bartlett and Snelus (1981) presented middle-aged and older listeners with music popular from various decades for a familiarity and time-last-heard judgment, and lyric recall for tunes deemed familiar. Of course, they could not guarantee that listeners had been exposed to all

the tunes, but the older listeners had a higher proportion of “familiar” judgments than the middle-aged adults, for music popular when the former group were young adults but the latter group were children or not yet born. This early-learned advantage was confirmed by Rubin et al. (1998) and also Schulkind et al. (1999), who showed that these songs elicit high emotionality ratings, which could partially explain the memory advantage.

The retention of familiar music over decades may depend on the extent to which the music was the focus of attention during early exposure. Maylor (1991) found that older adults were worse than middle-aged adults in recognizing television themes no matter what the retention interval (i.e., era of learning). However, it could be the case that this kind of incidental music has less musical and emotive meaning than music learned among peers as a young adult, or in settings such as summer camp or as part of religious services. It is possible that older adults have a particular disadvantage in the very casual learning situations of hearing background music to a television show.

Overall, research seems consistent with everyday observations that older adults can store representations of music for decades, but does this memory ability extend to newly learned music? As noted earlier, there is some folk belief that memory for music may be somewhat protected from the usual age-related impairments in episodic memory. However, it seems that at least in episodic memory over the long term, aging is associated with the kinds of impairments that are seen in other domains. Again, research is sparse but there are a few such studies.

Two relevant studies were already described in the context of experience and tune-knowledge effects. One was the Halpern et al. (1995, Experiment 2) study, which presented 24 unfamiliar melodies, each four times in three different keys, for ratings on pleasantness. In a subsequent old–new recognition test, musical experience did not affect performance (the point made earlier), but young adults were significantly (but not drastically) better than older adults. In the second relevant study, Bartlett et al. (1995) presented well-known and novel tunes (the latter permutations of the well-known tunes) for old–new recognition. In Experiment 2, the tunes were presented blocked by knowledge (i.e., only well-known tunes in one study list and test, only novel tunes in another), whereas in Experiment 3, well-known and novel tunes were mixed in each study list and test. Young adults performed better than older adults in both of the experiments particularly because older adults had large false alarm rates to familiar tunes. The size of the age difference was stronger in the mixed condition in which both the young and old began having trouble suppressing false alarms to new but well-known items (a problem attributed to familiarity in the absence of naming). Blanchet et al. (2006) found that older adults had hit rates equivalent to younger adults when asked to memorize a short set of unfamiliar tunes for later recognition, but had trouble suppressing false alarms. An encoding task (classifying the tunes as march or a waltz) actually hurt older people’s performance. The authors suggested that the task did not provide enough distinctive cues but served instead as a divided attention task, consistent with the earlier point about the lack of distinctiveness in memory encoding.

From the small amount of evidence available, it seems that retention of a set of items for recognition later in the experimental session is subject to the same age-related decline as seen in many other domains (Park and Schwarz 2000). However, short-term retention in same–different tests does not always show an age-related deficit. Meinz (2000) presented musical notation in a variety of short-term memory tasks, both recall and recognition, to musically literate people of various ages. She found age-related deficits in only a few memory tasks, and no age-related impairment in her composite memory score. Halpern et al. (1995) found that older adults were less adept than younger adults in differentiating exact transpositions from different-contour transpositions. But it was this study that found age invariance when the task was to differentiate exact transpositions from changed-interval transpositions. So the age-related deficit here seemed more tied to the more global task of contour processing, not the detailed task of interval detection. All listeners performed the contour task much more accurately than the interval task, belying another folk belief that age-related impairments are always larger in harder tasks.

Only small deficits due to aging were found in another type of short-term comparison task (Dowling et al. 2008). This was task mentioned earlier of comparing familiar or unfamiliar standards with an exact repetition or a changed-interval target, at slow, medium, or fast speeds. Whereas results showed a large effect of experience, only a modest effect of age occurred, even for tunes going very fast or very slow. This is surprising because one reasonable hypothesis was that older adults might have trouble integrating a very fast stream of notes due to attentional problems, or remembering a very slow stream of notes due to working memory limitations. But these near side-by-side comparisons do not task the more deliberate encoding used in list learning experiments and where deficits are the hallmark of cognitive aging.

One aspect of memory not so far addressed is implicit testing. Most of the studies presented so far involve explicit testing, usually recognition. Implicit tests involve a change in behavior without the testee necessarily experiencing a conscious memory act. And in fact, in real life people often retrieve music without necessarily having a memory retrieval experience. For instance, a person may hum along with a song on the radio without being able to recall the tune by name, or find that she or he likes a tune for some reason, only later realizing it had been heard previously. A friend told an anecdote of suddenly feeling sad while a certain hymn was being sung at her church. Only later did she remember that the tune had been sung at her mother's funeral.

A few studies have found that music can be tested implicitly. Warker and Halpern (2005) adapted a stem completion task to music: a list of unfamiliar tunes was presented, followed by the first few notes (stems) of old or new tunes. People were asked to hum a note that “sounded good” after the stem. They sang the correct note more often for old than new tunes, independent of explicit memory for that note. Peretz et al. (1998) found dissociations of recognition memory for music (explicit) from increases in liking for old tunes (mere exposure effect, implicit).

A common finding in cognitive aging is that implicit testing often reveals smaller aging effects than explicit testing (Fleischman et al. 2004), possibly due to more automatic nature of the encoding and/or retrieval processes used in implicit tests compared to explicit. Is this result also shown in studies with music? Gaudreau and Peretz (1999), and Halpern and O'Connor (2000), showed that recognition memory was quite impaired in older versus younger adults, but age made no difference in the implicit task of an affective judgment to each tune. Thus it may be the case that effortful retrieval is a locus of age-related effects than encoding, as the tunes had to be encoded to be rated as more pleasant or better liked. On the other hand, it might be argued that elaborative encoding at the time of study is important for explicit-test performance but not for implicit-test performance, and that older persons are deficient at such encoding. However, it was argued earlier that elaborative encoding strategies seem largely ineffective in changing music recognition performance, and thus is evidence against this view.

The final question raised in this section is whether the deleterious effects of normal aging and the beneficial effects of experience can offset one another. That is, are there any situations in which age and experience interact? It turns out that this is a perhaps desired but not-often-found pattern in various domains. For instance, Morrow et al. (1994) failed to find this for airline pilots, except for one or two specific tasks. Mainz (2000) did not find age by experience interactions in her notation memory studies. A review of work from the research program of the current authors (Halpern and Bartlett 2002) examined 13 experiments that could have revealed such a compensatory pattern. In only one instance did such a pattern obtain, and even there, the interaction accounted for very little variance. In fact, that review concluded that age and experience typically affected different tasks and that the benefit of younger age and more experience are not interchangeable.

The opposite side of this coin is that nothing suggests that age diminishes the positive effect of experience. In fact, Mainz (2000) found that because experience usually increases with age, this “confound” can lead to an apparent diminution of age effects with experience. So even though this is not an interaction in the theoretical sense, in a practical sense, older musicians would be expected to exceed younger nonmusicians, in experience-sensitive tasks.

8.7 Conclusions and New Directions

Perhaps the major message of this chapter is that memory for melodies depends upon knowledge. First, it depends on knowledge of individual tunes, their perceived familiarity and nameability. Second, it depends on knowledge of the tonal structure and well-formedness of tunes, including knowledge of in-key versus out-of-key notes. Finally, it depends on the musical knowledge of the listener, using the term rather broadly to include both symbolic knowledge and procedural skills developed in the course of musical training.

Knowledge of individual tunes is important in the simple short-term memory same–different task of judging pairs of tunes as same or different. Performance is much higher if the first-presented tune in a pair is a well-known melody, and this is true whether or not the task requires recognition of targets that have been transposed to different keys, so long as ceiling effects are avoided and accurate same–different judgments cannot be based on global aspects of the tunes such as melodic contour. In the domain of long-term memory, tunes that have been rated as highly familiar are recognized more quickly (i.e., after fewer notes) than those rated as only moderately familiar, and prior knowledge of tunes is also important for “episodic memory,” that is, recollecting the contexts in which tunes have been presented. Recollection of context appears to depend not simply on a tune being familiar to the listener, but on its nameability; that is, its unique identifiability with a proper name, word, or phrase. It is unknown whether familiarity without nameability is sufficient to produce (1) quicker identification of tunes and (2) high performance in short-term same–different tasks, including those requiring accurate encoding of musical interval information. Regarding the latter point, it is clear that listeners have rather poor knowledge of the intervals of novel tunes heard only once before, and yet these same listeners – even if they are musically untrained – have accurate knowledge of the intervals of tunes they know well. An open question is whether accurate knowledge of the intervals of well-known tunes can be developed with tunes that have been heard repeatedly without ever being linked to names or other verbalizable information that uniquely identifies them.

It will surprise no one that tunes that conform to familiar tonal structures are easier to recognize, and indeed they are. However, researchers have only started to examine what aspects of tonal structure in a given musical culture are important for learning and remembering of melodies, and how these aspects of tonal structure themselves are learned. The research covered here suggests that a very basic aspect of tonal structure – the set of in-key versus out-of-key notes – is implicitly learned by virtually all listeners, and produces effects of tonal structure on memory. However, it is unclear whether other aspects of tonal structure might affect melodic memory at different levels of musical expertise. In light of evidence that nonmusicians show relatively poor differentiation in their ratings of the centrality of different notes within a key (Krumhansl 1990), and are poor at classifying melodies as major versus minor (Leaver and Halpern 2004), it is likely that such more subtle aspects of tonal structure will affect melodic memory only among the more highly experienced or trained. It likewise will surprise no one that more musically trained listeners show better memory for tunes. However, the research in this area has advanced to a point quite beyond common knowledge. Certainly, few “people on the street” would intuit that musicians do not differ from nonmusicians in their sensitivity to a melody’s out-of-key note, but that musicians are better in making the basic judgment of whether one novel melody is a transposition of one heard a few seconds before (this task is trivial even for nonmusicians if the first melody is known, but is more difficult if the first melody is novel).

The ability to group musical materials during encoding may play a role in the effects of expertise. In chess, for example, it is very well known that experts chunk

together multipiece configurations of chess pieces, enjoying very high memory for chess-board displays as a consequence (Chase and Simon 1973). Similarly, much research with faces – with which it is argued all of us are experts – supports this chunking, or configural, encoding. Moreover, recent evidence suggests similar forms of configural encoding can emerge with expertise in identification of birds, automobiles, and invented three-dimensional forms (i.e., “greebles”; see Bukach et al. 2006). In fact, research summarized by Bukach et al. suggests that two different types of configural processing – holistic processing of the whole object and relational processing of spatial relations among features – both are related to expertise with visual stimuli. Although these two subtypes of configural processing are separable in terms of brain function (and probably in other ways), they may be functionally related in that attention to a whole object is likely to facilitate encoding of spatial relations among its constituent features.

By analogy, attention to the whole of a musical phrase might impair selective processing of individual notes, but improve the encoding of musical relations among these notes. Indeed, the Dalla Bella et al. (2003) study cited earlier in this chapter, as well as research by Schulkind and colleagues (Schulkind 2004; Schulkind et al. 2003), has shown that the most important notes for identifying melodies tend to occur at boundaries of musical phrases of five to seven notes. In another relevant study, Kim and Levitin (2002) replaced the notes of well-known melodies with bandpass filtered sounds that severely disrupted the absolute and relative pitch of the individual notes. Melody identification was approximately 75% when bandpass filtering had reduced identification of individual pitches and inter-note intervals to almost 0%, a striking example of tune recognition based on inter-note relations when the notes themselves are not accurately encoded. Unfortunately, it is not yet clear whether the relational processing supported by these studies is linked to expertise. Dalla Bella et al. (2003) found that both musicians and nonmusicians appeared to recognize melodies through processing of phrase-level units, and Schulkind (2004) found no reliable correlations between musical training and inter-condition differences that would have suggested a linkage of phrase-level coding to musical expertise. Finally, Kim and Levitin’s listeners had at least 10 years of musical training, raising the question of whether untrained individuals would show similar evidence for relational recognition of melodies – or not.

One of the most encouraging findings pertaining to musical expertise is that its enhancing effects on melodic processing holds up well in old age. Although age-related deficits in melodic processing have been found, the melodic processing advantages linked to music training appear not to decline at all in old age. Moreover, age-related deficits in melodic processing do not appear to involve the more intrinsically musical aspects of melodies such as tonality, key, or chroma. Thus, so far the evidence is quite well aligned with anecdotal reports of preserved memory for music among the very old and demented

The effects of expertise have been examined primarily in short-term memory tasks, and, of all the many gaps in the literature to date, perhaps none is more striking than the lack of information about expertise effects in long-term melodic memory. It certainly is possible that such expertise effects are present, and remain largely

unknown simply because they have not been examined. On the other hand, the processes and representations used in long-term melodic memory may be fundamentally different than those used in short-term memory tasks. Kosslyn's (1980) influential theory of visual imagery drew a sharp distinction between the "surface display" underlying the experience of visualizing an object and the "deep representations" that support long-term retention of visual information (the latter being at least partly propositional). Research and theory on melodic processing, and on how musical knowledge affects such processing, should be directed at this question.

Acknowledgments We thank W. Jay Dowling for many helpful suggestions during the preparation of this chapter and Kay Ocker for help in preparation of the manuscript.

References

- Andrews MW, Dowling WJ, Halpern AR, Bartlett JC (1998) Identification of speeded and slowed familiar melodies by younger, middle-aged, and older musicians and nonmusicians. *Psychol Aging* 13:462–471.
- Bartlett JC, Dowling WJ (1980) The recognition of transposed melodies: a key-distance effect in developmental perspective. *J Exp Psychol Human* 6:501–515.
- Bartlett JC, Snelus (1981) Lifespan memory for popular songs. *Am J Psychol* 93:551–560.
- Bartlett JC, Halpern AR, Dowling WJ (1995) Recognition of familiar and unfamiliar music in normal aging and Alzheimer's disease. *Mem Cognition* 23:531–546.
- Benjamin AS (2008) Memory is more than just remembering: strategic control of encoding, accessing memory, and making decisions. In Benjamin AS, Ross BH (eds), *Skill and Strategy in Memory Use*. Amsterdam: Elsevier, pp. 175–223.
- Blanchet S, Belleville S, Peretz I (2006) Episodic encoding in normal aging: attentional resources hypothesis extended to musical material. *Aging Neuropsychol C* 13:490–502.
- Boltz M, Marshburn E, Jones MR, Johnson WW (1985) Serial pattern structure and temporal order recognition. *Percept Psychophys* 37:209–217.
- Brattico E, Tervaniemi M, Näätänen R, Peretz I (2006) Musical scale properties are automatically processed in the human auditory cortex. *Brain Res* 1117:162–174.
- Bukach CM, Gauthier I, Tarr MJ (2006) Beyond faces and modularity: the power of an expertise framework. *Trends Cogn Sci* 10:159–166.
- Chase WG, Simon HA (1973) The mind's eye in chess. In Chase WG (ed), *Visual Information Processing*. New York: Academic Press, pp. 215–281.
- Craik FIM, Lockhart RS (1972) Levels of processing: a framework for memory research. *J Verb Learn Verb Be* 11:671–684.
- Cross I, Howell P, West R (1983) Preferences for scale structure in melodic sequences. *J Exp Psychol Human* 9:444–460.
- Cuddy LL, Lyons HI (1981) Musical pattern recognition: a comparison of listening to and studying tonal structures and tonal ambiguities. *Psychomusicology* 1:15–33.
- Dalla Bella S, Peretz I, Aronoff N (2003) Time course of melody recognition: a gating paradigm study. *Percept Psychophys* 65:1019–1028.
- Deakin JM, Allard F (1991) Skilled memory in expert figure skaters. *Mem Cognition* 19:79–86.
- Demorest SM, Morrison SJ, Beken MB, Jungbluth D (2008) Lost in translation: an enculturation effect in music memory performance. *Music Percept* 25:213–223.
- Deutsch D (1979) Octave generalization and the consolidation of melodic information. *Can J Psychology* 33:201–205.

- Dewitt LA, Crowder RG (1986) Recognition of novel melodies after brief delays. *Music Percept* 3:259–274.
- Dowling WJ (1978) Scale and contour: two components of a theory of memory for melodies. *Psychol Rev* 85:341–354.
- Dowling WJ (1982) Chroma and interval in melody recognition: effects of acquiring a tonal schema. *J Acoust Soc Am* 72:S11 (Abstr).
- Dowling WJ (1991) Pitch structure. In Howell P, West R, Cross I (eds), *Representing Musical Structure*. London: Academic Press, pp. 33–57.
- 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*. New York: Academic Press.
- Dowling WJ, Kwak S-Y, Andrews MW (1995) The time course of recognition of novel melodies. *Percept Psychophys* 57:136–149.
- Dowling WJ, Tillman B, Ayers DF (2002) Memory and the experience of hearing music. *Music Percept* 19:249–276.
- Dowling WJ, Bartlett JC, Halpern AR, Andrews MW (2008) Melody recognition at fast and slow tempos: effects of age, experience, and familiarity. *Percept Psychophys* 70:496–502.
- Eysenck MW (1979) Depth, elaboration and distinctiveness. In Cermak LS, Craik FIM (eds), *Levels of Processing in Human Memory*. Hillsdale, NJ: Lawrence Erlbaum, pp. 89–118.
- Fleischman DA, Wilson RS, Gabrieli JDE, Bienias JL, Bennett DA (2004) A longitudinal study of implicit and explicit memory in old persons. *Psychol Aging* 19:617–625.
- Gardiner JM, Radomski E (1999) Awareness of recognition memory for Polish and English folk songs in Polish and English Folk. *Memory* 7:461–470.
- Gaudreau D, Peretz I (1999) Implicit and explicit memory for music in old and young adults. *Brain Cognition* 40:126–129.
- Glanzer M, Adams JK (1985) The mirror effect in recognition memory. *Mem Cognition* 13:8–20.
- Halpern AR (1988) Perceived and imagined tempos of familiar songs. *Music Percept* 6:193–202.
- Halpern AR (1989) Memory for the absolute pitch of familiar songs. *Mem Cognition* 17:572–581.
- Halpern AR, Bartlett JC (2002) Aging and memory for music: a review. *Psychomusicology* 18:10–27.
- Halpern AR, Müllensiefen D (2008) Effects of timbre and tempo change on memory for music. *Q J Exp Psychol* 61:1371–1384.
- Halpern AR, O'Connor MG (2000) Implicit memory for music in Alzheimer's disease. *Neuropsychology* 14:391–397.
- Halpern AR, Bartlett JC, Dowling WJ (1995) Aging and experience in the recognition of musical transpositions. *Psychol Aging* 10:325–342.
- Hébert S, Peretz I (1997) Recognition of music in long-term memory: are melodic and temporal patterns equal partners? *Mem Cognition* 25:518–533.
- Idson WL, Massaro DW (1978) A bidimensional model of pitch in the recognition of melodies. *Percept Psychophys* 24:551–565.
- Jones MR, Ralston JT (1991) Some influences of accent structure on melody recognition. *Mem Cognition* 19:8–20.
- Jones MR, Summerell L, Marshburn E (1987) Recognizing melodies: a dynamic interpretation. *Q J Exp Psychol* 39A:89–121.
- Kim J-K, Levitin DJ (2002) Configural processing in melody recognition. *Canadian Acoustics* 30:156–157.
- Korenman LM, Peynircioğlu ZF (2004) The role of familiarity in episodic memory and metamemory for music. *J Exp Psychol Learn* 30:917–922.
- Kosslyn SM (1980) *Image and Mind*. Cambridge, MA: Harvard University Press.
- Krumhansl CL (1990) *Cognitive Foundations of Musical Pitch*. Oxford: Oxford University Press.
- Leaver AM, Halpern AR (2004) Effects of training and melodic features on mode perception. *Music Percept* 22:117–143.

- Levitin DJ (1994) Absolute memory for musical pitch: evidence from the production of learned melodies. *Percept Psychophys* 56:414–423.
- Levitin DJ, Cook PR (1996) Memory for musical tempo: additional evidence that auditory memory is absolute. *Percept Psychophys* 58:927–935.
- Lucas BL, Schubert E, Halpern AR (2010) Perception of emotion in sounded and imagined music. *Music Percept* 27:399–412.
- Lynch MP, Eilers RE (1992) A study of perceptual development for musical tuning. *Percept Psychophys* 52:599–608.
- Maylor EA (1991) Recognizing and naming tunes: memory impairment in the elderly. *J Gerontol* 46:P207–217.
- McAuley JD, Stevens C, Humphreys MS (2004) Play it again: did this melody occur more frequently or was it heard more recently? The role of stimulus familiarity in episodic recognition of music. *Acta Psychol* 116:93–108.
- Meinz EJ (2000) Experience-based attenuation of age-related differences in music cognition tasks. *Psychol Aging* 15:297–312.
- Mikumo M (1992) Encoding strategies for tonal and atonal melodies. *Music Percept* 10:73–82.
- Morrow DG, Leirer VO, Alteiri PA, Fitzsimmons C (1994) When expertise reduces age differences in performance. *Psychol Aging* 9:134–148.
- Mungan E, Peynircioglu Z, Halpern AR (submitted) The effects of orienting task and familiarity on remembering and knowing in melody recognition.
- Park D, Schwarz N (eds) (2000) *Cognitive Aging: A Primer*. Philadelphia, PA: Psychology Press.
- Patel AD (2008) *Music, Language and the Brain*. New York: Oxford University Press.
- Peretz I, Gaudreau D, Bonnel A-M (1998) Exposure effects on music preference and recognition. *Mem Cognition* 26:884–902.
- Radvansky GA, Fleming KJ, Simmons JA (1995) Timbre reliance in nonmusicians' and musicians' memory for melodies. *Music Percept* 13:127–140.
- Rubin DC, Rahal TA, Poon LW (1998) Things learned in early adulthood are remembered best. *Mem Cognition* 26:3–19.
- Schacter DL, Tulving E (1994) What are the memory systems of 1994? In Schacter DL, Tulving E (eds), *Memory Systems 1994*. Cambridge, MA: The MIT Press, pp. 1–38.
- Schellenberg EG, Trehub SE (2003) Good pitch memory is widespread. *Psychol Sci* 14:262–266.
- Schulkind MD (1999) Long-term memory for temporal structure: evidence from the identification of well known and novel songs. *Mem Cognition* 27:896–906.
- Schulkind MD (2004) Serial processing in melody identification and the organization of musical semantic memory. *Percept Psychophys* 66:1351–1362.
- Schulkind MD, Hennis LK, Rubin DC (1999) Music, emotion, and autobiographical memory: they're playing your song. *Mem Cognition* 27:948–955.
- Schulkind MD, Posner RJ, Rubin DC (2003) Musical features that facilitate melody identification: how do you know it's "your" Song when they finally play it? *Music Percept* 21:217–249.
- Shepard RN, Jordan DS (1984) Auditory illusions demonstrating that tones are assimilated to an internalized musical scale. *Science* 226:1333–1334.
- Standing L (1973) Learning 10,000 pictures. *Q J Exp Psychol* 25:207–222.
- Tulving E (1983) *Elements of Episodic Memory*. New York: Oxford University Press.
- Warker JA, Halpern AR (2005) Musical stem completion: humming that note. *Am J Psychol* 118:567–585.
- Warren RM, Gerdner DA, Brubaker BS, Bashford JA (1991) Melodic and nonmelodic sequences of tones: effects of duration on perception. *Music Percept* 8:277–290.
- Wright AA, Rivera JR, Hulse SH, Shyan M, Neiwirth JJ (2000) Music perception and octave generalization in rhesus monkeys. *J Exp Psychol Gen* 129:291–307.
- Yonelinas AP (2002) The nature of recollection and familiarity: a review of 30 years of research. *J Mem Lang* 46:441–517.

Index

A

- Accents, beat, meter, 168, 186ff
 - melodic and temporal, 187ff
- Acoustic cues, perception of emotion in music, 107ff
- Acoustic properties, pulse-resonance sounds, 18–20
- Acoustic scale, effects on pitch perception, 39ff
 - pitch definition, 44–45
 - timbre definition, 44–45
 - time-interval profile, 36
- Acquisition of music, 89ff
- Acquisition, rhythm, 99ff
- Aesthetic emotions, 130
- Affective responses, music, 129ff
- Aging effects, memory for melody, 249ff
- Alzheimer's disease, memory for melody, 250
- Animals, pulse-resonance sounds, 16–17
- Appraisal theory, emotion, 131
- Arousal potential, music, 152
- Arousal, valence, 135
- Atonal music, see Nontonal Music
- Attention, Dynamic Attending Theory (DAT), 171, 187
 - effects of musical training, 117–118
 - memory for melody, 251
- Auditory image, 30ff
 - cochlea, 32–33
 - model, 31ff
 - neural activity pattern, 33–35
 - outer and inner ear processing, 32
 - spectral profile, 35–35
- Auditory perception, acoustic properties, 37ff
 - time-domain model, 30ff
- Auditory system, representation of musical tones, 30ff

B

- Bach, 65–66
- Bali, music, 67
- Beat and meter, 167–168, 186ff
- Bifurcation analysis, 204
- Brain activation, emotion, 142ff
- Brain damage, musical emotion, 5–6, 144
- Brain imaging, rhythm, 223
- Brain processing of music, musicians vs. nonmusicians, 95–97
- Brain processing, children and pitch, 97–98
- Brain, music processing, 94ff
 - rhythm processing, 103–104
- Brass instruments, excitation, 24–25
- Bursting, neural systems, 223–224

C

- Canonical model, neural dynamics, 204ff
- Categorical models, emotion, 130
- Children, cultural cues to emotion in music, 110–111
 - experience and perception of emotion in music, 111–112
 - musical skill acquisition, 89ff
 - musical training and rhythm processing, 104–105
 - response to emotion in music, 109–110
 - rhythm processing, 100–101
 - tonal hierarchy, 60–62
 - tone changes with body size, 18–19
 - training and pitch perception, 97–98
- Chills, emotion, 147ff
- Circumplex model, model of emotion, 130, 134
- Cochlea, auditory image, 32–33
- Cochlear resonance, pitch perception, 213ff
- Cognitive reference points, tonal hierarchy, 53
- Cognitive science, tonal hierarchy theory, 79ff

- Cognitive skills, effects of musical training, 113ff
- Cognitivists, 131ff
- Complex meters, 168, 187–188
- Computational models, tonal hierarchy, 72–74
- Consonance and dissonance, 150, 217
neuroimaging, 143–144
- Contour information, memory for melody, 243
- Cortex, music processing, 94ff
representation of melodies, 92–93
- Cross-cultural studies,
tonal hierarchy, 66–68
- Cultures, and melody memory, 244–245
cues to emotion in music, 110–111
responses to rhythm, 102
- D**
- Detuning, oscillators, 211
- Development of musical perception,
enculturation, 99ff
- Development, music perception, 2, 3, 99ff
pitch organization, 90ff
tonal hierarchy, 60–62
- Dissonance and consonance, 150
- Driving rhythm, 170
- Dynamic attending theory (DAT),
entrainment, 170ff, 187
- Dynamic shape, memory for melody, 243
- E**
- Ear processing, auditory image, 32
- Eastern music, 139
- EEG and EMG, emotion, 142
- EEG, and rhythm, 223–224
- Elaborative encoding, memory for
melody, 238–240
- Emotion in music, cultural cues, 110–111
- Emotion perception in music, 106ff
- Emotion, development, 139
music, 3, 129ff
structure of, 134ff
- Emotivists, 131ff
- Enculturation, development of musical
perception, 99ff
- Enculturation, responses to rhythm, 102
- Entrainment theories, rhythm, 170ff
- Entrainment, Dynamic Attending Theory
(DAT), 170ff
rhythmic neural bursting, 225
spontaneous oscillation, 208–209
- Episodic memory, melody, 239–240
- Evaluative space model, emotion, 136
- Excitation, brass instruments, 24–25
woodwinds, 24–25
- Experience and knowledge, role in perception
of structure, 189
- Experience, development of music skills, 3
effects on perception of emotion
in music, 111–112
- F**
- Familiarity, and emotion, 150ff
in absence of tune recognition, 237
melody, 235–236
- Figural coding, rhythm, 166
- Filter, music al instruments, 25–27
- Filtering, dimension of timbre, 45
- fMRI, music processing, 94–95
- G**
- Glottal pulse rate (GPR), 19–20
- Glottal pulse resonator, human body size,
23–24
- Grouping effects, rhythm, 166–167
- Grouping, rhythm, 183ff
- H**
- Harmonic structures, 90
- Harmony, sensitivity to, 93
- Hebbian learning, neural dynamics, 211ff, 220
- Helmholtz, 226
- Higher-order resonance, 209–210
- History, music research, 4–5
- Hodgkin-Huxley models, 202, 204
- Hopf bifurcation, nonlinear oscillation, 206–208
- Human body size, and vocal tone
changes, 18–19
- Human size, glottal pulse resonator, 23–24
- Human voice, source, 23–24
- Humans, pulse-resonance sounds, 17–18
- I**
- Imagination-Tension-Prediction-Response-
Appraisal (ITPRA), theory, 132
- Indian ragas, comparison to
Western music, 66–67
- Individual differences, cognition
and attention, 113–118
musicians vs. non-musicians, 94ff
tonal hierarchy, 63–64
- Infants, acoustic cues for perception of
emotion in music, 108–109

- perception of emotion in music, 107
 - pitch acquisition, 89ff
 - rhythm processing, 100–101
- Instrument families, pitch, 14–16
- Instrument making, 27ff
- Instrument register, 14–16
- Instrument size, tone changes, 18–19
- Instruments, see also Stringed Instruments,
 - Woodwinds, Brass
 - scaling, 26, 27ff
 - size, 26, 27ff
 - sustained-tone, 21ff
- Intelligence, effects of musical training, 116–117
- Internal clock, rhythm, 168–169
- Interval clock models, rhythm, 191
- Interval theories, rhythm, 169–170

- J**
- Joint accent structure (JAS), 187–188

- K**
- Key-finding models, tonal hierarchy, 74–76

- L**
- Language skills, effects of musical training, 115–116
- Learning, connectivity, 211–212
 - tonal hierarchy, 60–62
- Long-term memory, melody, 236ff
- Loudness, definition, 37–38
 - emotion, 138

- M**
- Magnitude spectrum, vowel, 18
- Making of musical instruments, 27ff
- Mathematical abilities, effects of musical training, 113–114
- Measure, beats, 167–168
- Melodic structures, 90
- Melody acquisition, cortex, 92–93
- Melody, determination in instruments, 25
 - effect of spectral envelope shape, 45–46
 - memory, 233
 - perception, 4, 8ff, 39ff, 42–44
 - recognition, 4
 - Memory for names vs. tunes, 239
 - Memory of melody, human development, 246–247
 - Memory, effects of musical training, 117–118
 - melody, 233
 - Metamemory, memory for melody, 248
 - Meter perception, 167–168
 - Methodologies, tonal relationships, 55ff
 - Metric coding, rhythm, 166
 - Metric hierarchy, 167–168
 - Metrical structure, role of tempo, 188–189
 - Moods, 130, 133
 - Motor tempo, 172
 - Mozart effect, 114, 134
 - MRI, music processing, 94–95
 - Multiple look model, tempo, 179ff
 - Music acquisition, effects of musical experience, 89ff
 - Music emotion, perception in infants and children, 108–109
 - vocal cues, 106ff
 - Music perception, 2–3, 14–15
 - brain imaging, 5–6
 - development, 2, 3
 - pitch and timbre, 46–57
 - Music processing, cortex, 94ff
 - Music research, historical perspectives, 4–5
 - Music training, tonal hierarchy, 62
 - Music, acquisition, infants and children, 89ff
 - Bali, 67
 - bitonal Western, 68–69
 - emotion, 3, 129ff
 - listeners response, 2
 - neurodynamics, 4, 201ff
 - nontonal Western, 68–69
 - perception of emotion, 106ff
 - tonal hierarchy, 51ff
 - tonal, 51ff, 89ff
 - Western, 65–66
 - Musical characteristics, emotion, 137ff
 - Musical complexity, 6–8
 - Musical emotion, acoustic cues, 107ff
 - Musical experience, effects on music acquisition, 89ff
 - memory for melody, 247ff
 - Musical instruments, filter, 25–27
 - perception of, 13ff
 - pulse-resonance sounds, 18, 21ff
 - timbre, 15–16
 - Musical perception, development, 99ff
 - future research, 4ff
 - Musical pitch processing, adult musicians vs. nonmusicians, 94ff
 - Musical tension, tonal hierarchy, 76ff
 - Musical time, 3–4
 - Musical tones, representation in auditory system, 30ff
 - Musical training, 89ff
 - and memory for melody, 245–246
 - effects on cognitive skills, 113ff

- Musical training (*cont.*)
 effects on intelligence, 115–117
 effects on mathematical abilities, 113–114
 effects on memory and attention, 117–118
 effects on reading, 115–116
 effects on rhythm processing in
 children, 104–105
 effects on spatial-temporal abilities, 114–115
 perception of emotion in music, 111–112
- Musical universal tonality, 216ff
- Musicians vs. nonmusicians (adult), musical
 pitch processing, 94ff
 rhythm processing, 103–104
- N**
- Nameability, melody, 235–236
- Neural activity pattern (NAP),
 auditory image, 33–35
- Neural function, music, 4
- Neural networks, tonal hierarchy, 73–74
- Neural oscillation, 203
- Neurodynamics, music, 4, 201ff
- Neuroimaging, emotion, 143–144
- Neurological basis, tonal hierarchy, 63–64
- Noise, definition, 21
- Nonlinear amplitude responses, 209
- Nonlinear oscillation, Hopf bifurcation, 207–208
- Nonlinear resonators, neural dynamics, 205ff
- Nontonal music, Stravinsky, 68
- Nontonal Western music, tonal hierarchy, 68–69
- Novel tone sets, tonal hierarchy, 69–71
- O**
- Orchestra, sustained-tone instruments, 21ff
- Oscillator networks, resonance, 205ff
- P**
- Perceived beat, 167
- Perceived emotion, vs. felt emotion, 145
- Perception of emotion in music, experience
 effects, 111–112
- Perception of grouping, rhythm, 183ff
- Perception, emotion in music, 107ff
 melody, 4, 8ff, 39ff, 42–44
 music, 2–3
 musical instruments, 13ff
 pitch, 14–15
 register, 15–16
 relative pitch, 51ff, 201ff
- Perceptual fluency/attribution model,
 emotion, 152
- Performance accuracy,
 memory for melody, 236ff
- Physiological measures, emotion, 141ff
- Pitch acquisition, children, 89ff
 infants and children, 89ff
 training, 89ff
- Pitch acquisition, Western music, 90ff
- Pitch centrality, 52
- Pitch height, emotion, 138
- Pitch interval information, memory for
 melody, 240–241
- Pitch organization, development, 90ff
- Pitch perception, 1–2, 14–15, 213ff
 children and training, 97–98
 effects of acoustic scale, 39ff
- Pitch, acoustic scale, 44–45
 definition, 37–38
 effects of source size, 38–39
 effects on grouping, 185
 instrument families, 14–16
 space model, 76–77
 usefulness in music perception, 46–47
- Pitch-shift of the residue, 214–215
- Preferred perceptual tempo (PPT), 174
- Preferred tempo, 172–173
 genetics, 175
- Probe tone method, tonal hierarchy, 55–57
- Produced tempo, 181ff
- Psychological principles, tonal hierarchy, 53–55
- Pulse-resonance sounds, 16–18
 acoustic properties, 18–20, 37ff
 animals, 16–17
 musical instruments, 18
 production, 16–18
- Pulse-resonance tones, musical instruments, 21ff
- R**
- Reading, effects of musical training, 115–116
- Register, perception, 15–16
- Resonance, neural systems, 201ff
- Rhythm, 183ff
 acquisition, 99ff
 definition, 166
 effects of training on children, 104–105
 resonance, 220ff
- Rhythm and tempo, 165ff
- Rhythm models, 189ff
- Rhythm processing, adult musicians vs. adult
 nonmusicians, 103–104
 brain, 103–104
 infants and children, 100–101
- Rhythmic complexity, emotion, 138
- Rule-based, rhythm models, 190

S

- Sad-sounding music, 153–154
- Scalar expectancy theory (SET), rhythm, 169
- Scale, *see* Acoustic scale
 - definition, 21
- Scaling, instruments, 26, 27ff
- Schoenberg, twelve-tone serialism, 69
- Second dimension of pitch hypothesis, 44–45
- Self-reports, emotion, 140–141
- Semantic memory, memory for melody, 250
- Sensory dissonance, Helmholtz, 217–218
- Short-term memory, melody, 240ff
 - vs.* long-term memory for melody, 243–244, 252
- Signal detection theory, memory for melody, 240–241
- Simple meters, 168
- Singing, vocal folds, 23
- Size, instruments, 26, 27ff
- Sounds, of musical instruments, 13ff
- Source size, effects on pitch and timbre, 38–39
- Source, definition, 20–21
 - human voice, 23–24
 - musical instruments, 22
 - stringed instruments, 23–24
 - vocal folds, 23–24
- Spatial-temporal abilities, effects of musical training, 114–115
- Spectral envelope shape,
 - effect on melody, 45–46
- Spectral profile, auditory image, 35–36
- Spectrotemporal receptive fields (STRF), 216
- Spontaneous motor tempo (SMT), 173–174
- Stability, in tonal hierarchies, 204
 - neural dynamics, 204
- Startle response, 135
- State space, 204
- Statistical distribution, of musical tones, 65–66
- Statistical learning, tonal hierarchy, 60–62
- Stravinsky, nontonal music, 68
- Stringed instruments, source, 23–24
- Sustained-tone instruments, orchestra, 21ff
- Syncopation, 222

T

- Tempo, 172ff
 - absolute memory for, 175ff
 - and metrical structure, 188–189
 - and rhythm, 165ff
 - definition, 166
 - development with age, 174–175
 - discrimination, 177ff
 - JND, 179
 - memory for melody, 241
- Temporal contrast, rhythm, 171
- Temporal fluctuation, rhythm, 222
- Timbre, 2
 - acoustic scale, 44–45
 - definition, 37–38
 - effects of acoustic scale, 39ff
 - effects of source size, 38–39
 - emotion, 138
 - musical instruments, 15–16
 - usefulness in music perception, 46–57
- Time-domain model, auditory perception, 30ff
- Time-interval profile, acoustic scale, 36
- Timing, in music, 3–4
- Tonal center, in children, 61
- Tonal hierarchy in music, 51ff
- Tonal hierarchy theory, 2
 - implications for cognitive science, 53, 79ff
- Tonal hierarchy, computational models, 72–74
 - cross-cultural, 66–68
 - definition, 52ff
 - development, 60–62
 - individual differences, 63–64
 - key-finding models, 74–76
 - learning, 60–62
 - musical tension, 76ff
 - neural networks, 73–74
 - neurological basis, 63–64
 - nontonal Western music, 68–69
 - novel tone sets, 69–71
 - principles, 58–59
 - probe tone method, 55–57
 - psychology principles, 53–55
 - stability of tones, 204, 219–220
 - training *vs.* no training, 62
 - Western music, 65–66
- Tonal relationships, experimental methodologies, 55ff
- Tonal theories, 2, 51ff, 201ff
- Tonality, 2
 - memory for melody, 244, 246
 - neurodynamics, 216ff
- Tone centrality, 52
- Tone changes, body size in humans, 18–19
 - instrument size, 18–19
- Tone distributions, frequency or duration, 71–72
 - Western music, 65–66
- Training, musical, 89ff
- Transposition effects, memory for melody, 241

Tune structure, memory for melody, 244ff
Turkish listeners, memory for melody, 246
Twelve-tone serialism, Schoenberg, 69

V

Valence, emotion, 130ff
Vocal cues, emotion in music, 106ff
Vocal folds, singing, 23
 source of sound, 23–24
Vowel, magnitude-spectrum, 18

W

Weber's law, tempo discrimination, 177ff, 182
Western music, comparison to Indian ragas,
 66–67
 pitch acquisition, 90ff
 tonal hierarchy, 65–66
 tone distributions, 65–66
Wing and Kristofferson (W&K) model,
 tempo, 181ff
Woodwinds, excitation, 24–25