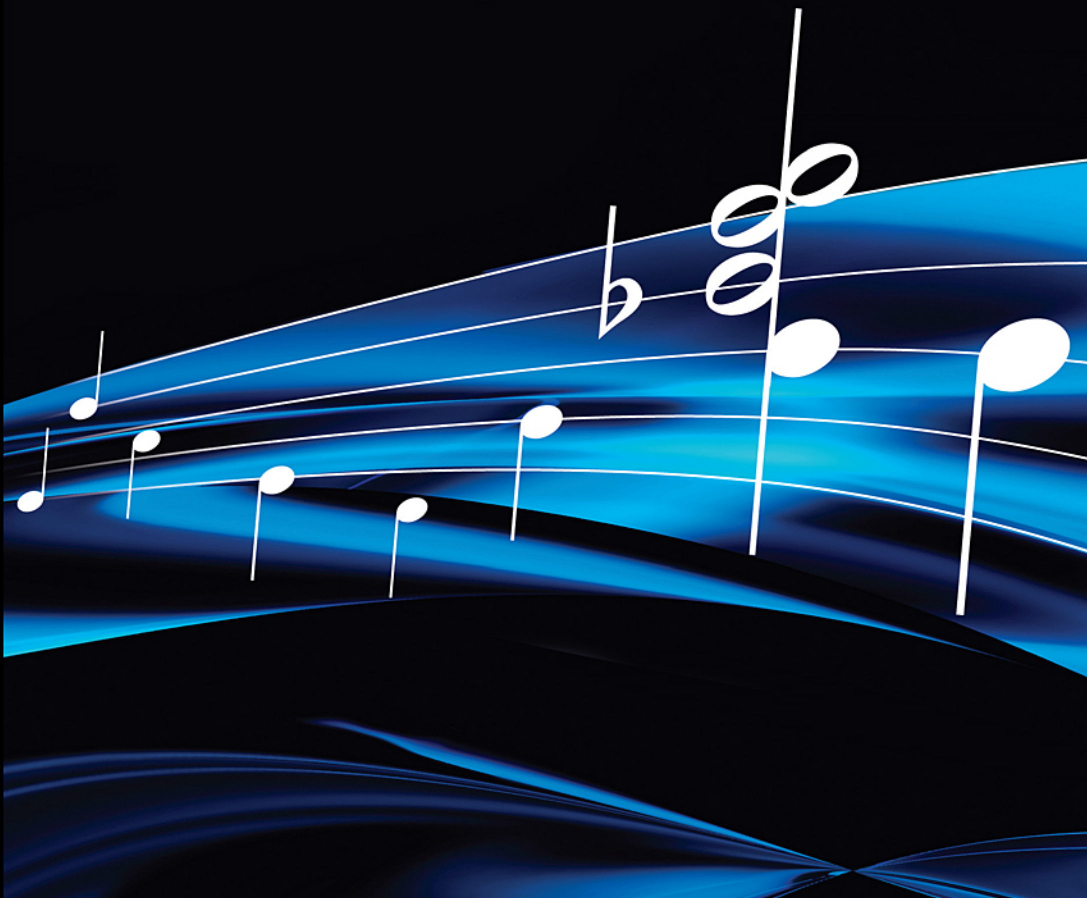
Psychology *of* Music

From Sound to Significance



Siu-Lan Tan, Peter Pfordresher and Rom Harré

Psychology of Music

Why are some disturbances of air molecules heard as 'noise' while others are perceived as music? What happens at the level of the sound wave, the ear, and the brain when we perform or listen to music? How do musical abilities emerge and develop, and become refined as one acquires musical expertise? And what gives music its deep emotional significance and its power to influence social behavior, across vastly different cultural contexts? These are some of the primary questions defining the field called 'the psychology of music' and driving the present volume.

This book provides an introduction to classic and current studies in the psychology of music, combining a comprehensive summary with critical assessments of existing research. The volume captures the interdisciplinary breadth of the field, while covering central topics in depth. Part I explores sound and music at an acoustic level, explaining auditory events with respect to the workings of the ear and brain. Part II focuses on perception and cognition of melody, rhythm, and formal structure. Part III examines the emergence and development of musical skills, and turns to the most practical aspects of psychology of music: music practice and performance. Finally, Part IV broadens the discussion to the question of meaning in music, with respect to its social, emotional, philosophical, and cultural significance. Throughout, both behavioral and neuroscientific perspectives are developed.

This book will be invaluable to undergraduate and postgraduate students in the fields of psychology and music, and will appeal to anyone else who is interested in the psychology of music.

Siu-Lan Tan is Associate Professor of Psychology at Kalamazoo College in Michigan, USA. She graduated in Music Education and Piano Pedagogy and taught music for many years in Hong Kong and California before completing an M.A. and Ph.D. in Psychology at Georgetown University. Her research has focused on musical unity, music notation, and film music, and her primary instrument is piano.

Peter Pfordresher is Associate Professor of Psychology at SUNY Buffalo in New York, USA. He completed his Ph.D. in Psychology at the Ohio State University. His research interests include the role of auditory feedback in music performance and causes of inaccurate singing. He has experience in musical performance with piano, trumpet, and voice, as well as song writing.

Rom Harré is Emeritus Fellow of Linacre College, Oxford University, UK. Currently he is Distinguished Research Professor of Psychology at Georgetown University, and Director of the Centre for the Philosophy of the Natural and Social Sciences at the London School of Economics. He has performed in various amateur ensembles on clarinet and saron.

Psychology of Music

From sound to significance

Siu-Lan Tan, Peter Pfordresher and Rom Harré

Psychology Press
Taylor & Francis Group
HOVE AND NEW YORK

First published 2010
by Psychology Press
27 Church Road, Hove, East Sussex, BN3 2FA

Simultaneously published in the USA and Canada
by Psychology Press
270 Madison Avenue, New York, NY 10016

This edition published in the Taylor & Francis e-Library, 2010.

To purchase your own copy of this or any of Taylor & Francis or Routledge's collection of thousands of eBooks please go to www.eBookstore.tandf.co.uk.

Psychology Press is an imprint of the Taylor & Francis Group, an Informa business

This publication has been produced with paper manufactured to strict environmental standards and with pulp derived from sustainable forests.

British Library Cataloguing in Publication Data
A catalogue record for this book is available
from the British Library

Library of Congress Cataloging-in-Publication Data
Tan, Siu-Lan, 1963–
Psychology of music : from sound to significance / Siu-Lan Tan, Peter Pfordresher, and Rom Harré
p. cm.
Includes bibliographical references and index.
ISBN 978–1–84169–868–7 (hb)
1. Music—Psychological aspects. 2. Music—Acoustics and physics. 3. Music—Social aspects. I. Pfordresher, Peter, 1971– II. Harré, Rom. III, Title.
ML3830.T167 2010
781′.11–dc22
2009038200

ISBN 0-203-85536-1 Master e-book ISBN

ISBN: 978-1-84169-868-7 (hbk)

Let us keep open the connections whereby the human spirit may freely move between the arts and the sciences, and thus make more of each.

– Yehudi Menuhin (1916–1999)
Quoted with permission of Hon. Mr. Gerard Menuhin

Contents

Preface

The inspiration for this book came from the authors' many years of experience in teaching psychology of music in various places and institutions. Our experience in teaching psychology of music began in the 1990s, when two of us team-taught a course with a music colleague Philip Tacka. Since then, the terrain of this field has changed considerably. New technologies and theories in several areas have stimulated significant developments. These include greater attention to neuroscientific research into the processes relevant to many aspects of musical life. There has also been progress in research on performance as well as on the perception of sound as music. New dimensions of the links between music and language have been explored. The focus of attention, previously very much on the cognitive psychology of music, has broadened in several directions including social, developmental, and applied psychology. We sought to create a book that reflects these important changes in this rich and expanding field.

This book is organized around the progression of musical sound from a physical stimulus to meaningful experience. *Part I: Foundations* explores the physical basis of musical experience in sound through an introduction to acoustics and neuroscience, explaining auditory events with respect to the workings of the ear and the brain. *Part II: The Perception and Cognition of Music* focuses on the perception of melody, rhythm and musical structure. *Part III: Development, Learning, and Performance* examines the emergence of auditory perception and early musical development, and music practice and performance at more advanced levels. Finally, *Part IV: The Meaning and Significance of Music* broadens the discussion to consider the social, emotional, philosophical, and cultural significance of music.

The design of this book is amenable to a variety of purposes. It can be used as a course text and is equally suited for perusal by interested lay readers. Introductory courses may begin with the *Foundations* chapters which assume no prior knowledge of acoustics, auditory perception, or neuroscience, while more advanced readers may start at the music-language discussion at the end of chapter 4 or at *Part II*. Those without a background in music theory may consider omitting chapters 7 and 13. Throughout the book, we present behavioral methods and neuroscience research as different techniques to address common issues, rather than as two fundamentally different research endeavors. Thus, neuroscientific findings are discussed in sections within the chapters on melody, rhythm, practice, performance, and emotion in music. Depending on the emphasis of the course or specific interest of lay readers, this book can also be read without the chapter on neuroscience or clearly marked sections on neuroscientific findings in subsequent chapters, without disturbing the continuity of the narrative. Our book focuses on basic research but also includes discussion of practical and applied topics such as concert hall acoustics, early childhood music education, practice techniques, music and consumer behavior, and film music. In our own teaching, we have found that this breadth of coverage allows the selection of material to be used at different levels and for different audiences.

We intend our book for a wide range of readers, and have kept technical terminology for music and psychology to a minimum in order to ensure the readability of the text for a broad audience. Only a general familiarity with music and psychology is assumed. The only exception is chapter 7, which discusses prominent music theories in some detail, and therefore necessarily requires more technical knowledge than the other chapters. Our illustrative figures do not rely heavily on musical notation, and musical examples in text draw from music from many genres, both classical and popular. Our intended audience includes casual readers exploring a new field, those seeking a foundation for advanced study in this area, and individuals who may become music psychologists and will later contribute to this field of study themselves. All readers, we hope, will emerge with a deeper understanding of the *amazing* human phenomenon of the creation and appreciation of sound as music.

Acknowledgments

Given the scope of this book, the individuals who contributed so generously of their time and expertise represent many different fields. We would like to express our appreciation to Steven Brown, Simone Dalla Bella, Paul Jeffries, Mari Riess Jones, Martin Lotze, Dirk-Jan Povel, David Temperley, and anonymous reviewers for their valuable suggestions on earlier drafts. We are also indebted to Ginevra Castellano, Roger Chaffin, Jennifer Cox, Patrik Juslin, Timothy Justus, Rohan Krishnamurthy, Scott Lipscomb, Josh McDermott, Rod Nave, Jessica Phillips-Silver, Christopher Plack, Michael Schutz, Keith Swanwick, Michael Tenzer, Laurel Trainor, and Robert Zivadinov for their many generous contributions. Extensive student input shaped this book from its earliest drafts to the final manuscript. Our thanks to Elizabeth Wakefield, Sally Warner, Christy Peaslee, Shanti Virupannavar, James Mantell, John Kulpa, Lauren Mitchell, Christina Violante, Amanda McFarland, Jackie Howard, Ryan Coppolo, Timothy Griffiths, Emily Dayton, and our vibrant psychology of music classes. We are especially grateful to the warm and wonderful staff at Psychology Press: To Lucy Kennedy and Tara Stebnicky for supporting the project, and to Sharla Plant, Nicola Ravenscroft, and Dawn Harris for their expert guidance. Thanks also to our copy editor Paula Clarke. Above all, to our families and partners Danny, Lyn, and Hettie: our deepest gratitude.

Siu-Lan Tan
Peter Pfordresher
Rom Harré
Michigan, New York, Oxford, and Washington DC.

1 The scope of psychology of music

People everywhere and at all times for which we have records have picked out certain patterns of sound for particular attention. Some of these patterns are the stuff of what we call 'music.' What are the characteristics of the sound patterns we recognize as music? Why is it that these sound patterns have had a special significance for human beings?

All perceptible sounds begin with a propagation of energy into the environment. It may be a light breeze setting into motion a thousand fluttering leaves, the plucking of the strings of a harp, or the striking of a bass drum. What makes the particular dance of air molecules 'musical' in some instances, while other disturbances of air molecules seem to give rise to mere sounds? Or noise?

It is not always desirable to try to give a formal definition of the topic of a program of study. We will not try to answer the question 'What is music?' in a neat, short formula. However, there are paradigm cases of music which most people in a particular culture can recognize. We can start with some exemplary cases from our own culture about which most would agree. A symphonic performance by an orchestra is music. Rock concerts are music. Advertising jingles are music. Many hear church bells, ringing out a simple melody, as music. However, not all sound is music. Could we give a similar catalogue of sounds that everyone would agree are not music? Perhaps the noise of a road drill or the sound of a lawn mower would be sound patterns that are exemplary of nonmusic. The whine of a vacuum cleaner or the gurgling of a dishwasher might strike us as obvious cases of nonmusic. But what about the sound of waves breaking on the beach? The howl of a wolf? Or the song of a bird?

While the extremes seem clearly distinguishable, there is no sharp line to be drawn between music and sounds that are not music. Though we can fairly easily identify paradigm cases of music and nonmusic, there are many patterns of sound that are not easily classified as one or the other. Whether they are taken up as music or not depends in part on the context in which they occur. Debussy imitated the sounds of breaking waves in *La Mer*. Messiaen composed remarkable imitations of birdsong in *Des Canyons aux Étoiles*, while Respighi's score for *The Pines of Rome* included a recording of a real nightingale's song. The blasts of real car horns punctuate Gershwin's energetic orchestration in *An American in Paris*. And Malcolm Arnold's *Grand, Grand Overture* included three vacuum cleaners among the orchestral instruments! There is also John Cage's *4′ 33″*, a composition in which the performer does nothing for 4 minutes and 33 seconds to allow ambient and incidental sounds to define the composition. Here, the boundaries between sound and music, and the very definition of music in the absence of sound controlled by the composer or musician, are questioned.

All-encompassing definitions of music may be elusive. Nevertheless, despite the fuzzy boundaries of the domain of music, it seems that there are auditory phenomena which we can generally agree are music, and on which there has been agreement in many cultures and historical epochs. Music is produced and perceived by human beings. Performers must learn the necessary skills to create ordered sound in meaningful patterns. Through exposure or training, listeners must learn to perceive those features of ordered sound patterns as music. There is clearly room for a systematic study of all these skills, as diverse as they are. The merging of psychology and music leads the way to such examination, and opens avenues to numerous and diverse topics for study. Newcomers are often surprised at the breadth and scope of this expansive field.

The scope of the field

The psychology of music as it is practiced in the twenty-first century is concerned with a great many questions. Among other things, it is concerned with *the processes by which people perceive, respond to, and create music, and how they integrate it into their lives*. These topics range from the way in which the ear extracts the pitch of a tone, to the way in which music is used to express or transform moods. Though this field of study makes important use of cognitive psychology, it also draws on many other branches of psychology such as sensation and perception, neuropsychology, developmental psychology, social psychology, and applied fields such as educational psychology.

The perspectives of each of these domains within psychology have shaped this book to some degree, as evident in the overall plan: Our exploration begins with a consideration of the physical properties of a sound wave, the transmission of sounds to the ear, and the neural bases of the perception and cognition of music. We then examine more closely the perception and cognition of melody, rhythm, and musical structure. Next, we trace the emergence and development of auditory capacities and musical abilities, and the acquisition of musical expertise, culminating in musical performance. Finally, we consider the question of meaning in music, and the social, emotional, and universal significance of music.

The psychology of music attracts not only psychologists and musicians, but scholars and researchers from a wide range of other disciplines. The present volume is also informed by perspectives from fields such as acoustics, neuroscience, musicology, education, philosophy, and ethnomusicology. We provide a brief overview here.

More than 2000 years ago it was realized that the musical possibilities of sound as heard were shaped and constrained by the physical properties of sound waves as they interacted with the amazing powers of the ear. Pythagoras linked the weight of a vibrating object to pitch, while his followers extended his intuitions to include the vibrations of strings linking pitch to the length of the string, and harmonics to the simple numerical ratios of those lengths. *Acoustics* is the science of the production, propagation and reception of those vibrations in the air that are relevant to hearing in general and music in particular. But it is also more – as discussed in chapter 2, the way that musical instruments and the human voice shape the physical processes that reach the listener, as well as the properties of the venues where music is produced and enjoyed, are also parts of the science of acoustics.

With respect to *neuroscience*, recent times have seen a surge of interest in the neural underpinnings of human musicality, and the current volume considers the way that neural activity may constrain or enhance our experience of music and music making. In order for us to experience music, the brain must pick up and import physical patterns from the auditory signal. Our studies must include an introduction to auditory neuroscience, beginning with the anatomy and physiology of the ear as presented in chapter 3. Of course, music is a complex and often multimodal experience. That is, it is not limited to hearing but involves other senses, such as sight and movement. Thus we must extend our discussion of neuroscience beyond audition, which we do in chapter 4. Recent discoveries from the field of neuroscience of music are also presented in subsequent chapters on perception of melody and rhythm, music practice and performance, and emotion in music in chapters 5, 6, 10, 11, and 14 respectively.

Insights from the field of *musicology* (the study of the structure and history of music) also continue to be essential to psychology of music, as evident in chapters 5, 6, and 7 on perception of melody, rhythm, and musical structure, and in chapter 14 on the emotional power of music. For example, in studying the power of music to 'express' emotions and its capacity to 'induce' emotional states in the listener, musicologists have addressed intriguing questions such as: How do musical structures give rise to emotions in listeners? Further, how does music come to have 'meaning' for the listener? The latter question is explored in chapter 13. A study of the basic theory of a practice involves bringing at least some of the underlying presuppositions to light as explicit principles, and subjecting them to critical examination. This is one of the ways that *philosophical* analysis is also an indispensable part of the psychology of music, bringing out certain presuppositions in the practices by which psychologists try to reach an understanding of music as a human phenomenon.

Within psychology of music, developmental psychologists are concerned with the emergence and maturation of musical behaviors. How these emerging abilities may best be supported and refined is a question for *music education*, the focus of chapter 9. Musical performance requires the development of a set of highly elaborated skills, and a growing base of knowledge that allows for the sensitive interpretation of music. Before a composition can be performed it must be composed. This too requires a highly sophisticated cluster of skills. Then there is the creative undertaking of improvisation; Western jazz and performances of Indian classical movement are notable examples. Innovative music education methods assume that all children are musical, and intensively immerse young children in creative and sensitive engagement with music with the aim of laying the foundations for lifelong musicality.

Finally, music is differentiated into a wide variety of musical cultures across the world, and as such, an anthropological perspective highlighting the study of distinct human cultures and their 'musics' is also relevant to an understanding of the psychology of music. Finding out what is common to all musical cultures and what seems to be unique to each one throws a great deal of light on psychological questions. These studies help us to differentiate between cultural sources of the features of a particular musical repertoire and those which may derive from the biological bases of musical perception. The application of anthropology to music is *ethnomusicology*, the theme of our final chapter. In the concluding chapter, we consider music as it is represented in a variety of societies around the world. But we shall not pretend to have answered entirely satisfactorily the enduring questions of what music *is*, how exactly it is created and received, and how music brings about the powerful effects that it does.

Range of research methods

Although the present discussion by no means exhausts all the questions and topics subsumed by the study of psychology of music, the richness and expansiveness of this field should already be apparent. Psychology of music is also distinguished by the scientific research methods commonly employed to study musical phenomena of interest, which leads one to a certain kind of explanation for the phenomena under investigation. There are many different kinds of explanations developed in the natural sciences, accompanied by various methods, and psychologists of music make use of many of them.

Throughout this book we will encounter a great variety of empirical studies, as various methodologies are needed to explore the whole gamut of musical experience. For some purposes, experiments manipulating specific experiential or phenomenological variables in carefully controlled conditions are useful. A researcher may manipulate the tempo (or pace) of a song to determine if it alters listeners' interpretations of the emotion that is being expressed. For other purposes, the recording and analyzing of real-time phenomena of musical production and experience is appropriate, such as when identifying the steps a concert pianist takes to learn and memorize a complex piece of music. In some instances, physiological methods or brain-imaging techniques may be employed, for example to examine changes in the body or brain while listening to or performing music. Qualitative studies and naturalistic observation also play an important part in studying the psychology of music. The nature of some aspects of musicality is best captured in rich observational studies of spontaneous activities in natural settings, such as observations of crowd behavior at live concerts or of parents singing to infants in their homes. Sometimes we can usefully sum up the musical experiences of a great many people in a sweeping generalization. In other cases, we need to pay close attention to music as it is perceived, appreciated, performed or composed by an individual, if we are to reach a sufficiently detailed understanding of what is happening.

Since the days of the pioneering volumes in this field, including Carl Seashore's *The Psychology of Musical Talent* and *Psychology of Music* (1919, 1938), Paul Farnsworth's *The Social Psychology of Music* (1958), Geza Révész's *Introduction to the Psychology of Music* (1954), and even the methodologically progressive Robert Francès' *The Perception of Music* (1958/1988), a wide variety of methods have been applied to the study of psychology of music. It is a great mistake in building up a science to insist on the primacy of any one of the many methods of research that can be found across the universe of the sciences. Indeed, the psychology of music is fascinating in part because it draws on so many diverse bodies of knowledge and multiple modes of investigation! Only by adopting such breadth can we hope to understand how this remarkable human practice is possible.

Coda

This introductory chapter has mapped the rich interdisciplinary terrain of the present book. The project was undertaken by three authors who represented three generations of academics when their paths first crossed almost 20 years ago as a distinguished professor, a graduate student, and an undergraduate major. Brought together many years later by a shared passion for this field, we collaborated on this volume in consultation with our many enthusiastic students who attended our psychology of music classes over the last several years. Clearly, the scope of inquiry of this field is vast, and no book can cover all that psychology of music encompasses. Our aim is simply to introduce readers to a broad range of classic and current research, and to ignite curiosity in the many intriguing questions and topics in this exciting and expanding field.

Considering what music meant to him after 50 years of studying the science of music, Carl Seashore, one of the important pioneers in this field, wrote in the preface to his 1938 volume of *Psychology of Music*:

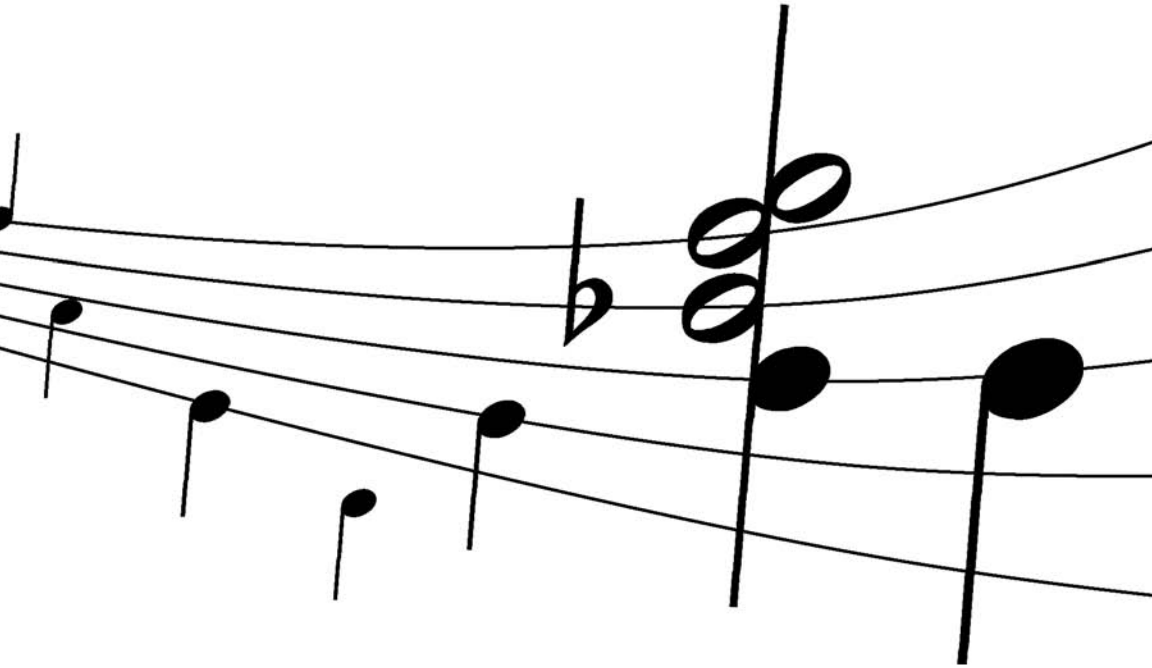
> Then I was a stargazer; now I am an astronomer. Then the youth felt the power of music and gave expression to this feeling in the way he loved

> and wondered at the stars before he had studied astronomy. Now the old man feels the same 'power of music,' but thinks of it in the manner that the astronomer thinks of the starry heavens. (p. xi)

Like Seashore, we hope this volume will serve to bridge the gap between the 'mere love and practice of music to an intelligent conception of it' (p. xi). At the same time, while learning the various principles and practices of astronomy, we hope our readers continue to be stargazers – moved by music and constantly fascinated by its many enduring mysteries.

Part I

Foundations



2 The acoustics of music

Brittle twigs and dry leaves crunching underfoot, rain pellets hammering down on a tin roof, the roar of a crowd at a football game. Each of these sounds has its characteristic acoustic features and patterns that make it distinct from others. A distant flute on a hillside, the shimmering tones of a *gamelan* ensemble in a garden courtyard, the thunderous *fortissimo* chord of a full symphonic orchestra in a grand concert hall. Each of these musical sounds, too, has a distinct acoustic signature. All these auditory events – musical and nonmusical – are dispersed into various sound environments: open air, outdoor arenas, and crowded concert halls. The sound waves often travel a great distance to our ears, and the physical patterns must be transformed into neural signals in the ear and brain. It sounds like a complex process. Yet most of us quite effortlessly recognize the sounds and their sources, defining some as 'noise' and some as 'music,' and describing them as 'sweet' or 'raucous,' 'mellow' or 'shrill.'

Acoustics is the science of sound; specifically, it is the study of the production, transmission, and reception of sound (*American Heritage Dictionary*, 2000, p. 15). The particular domain of *musical acoustics* focuses on the 'mechanisms of sound production by musical instruments, effects of reproduction processes or room design on musical sounds, [and] human perception of sound as music' (Hall, 2001, p. 2). Accordingly, the present discussion is organized around these three main topics. We will begin with a brief description of the *physical* characteristics of sound waves (*frequency*, *amplitude*, *complexity*), illustrate a few principles of acoustics as they apply to a variety of musical instruments, and finally explore the acoustics of some of the grand concert halls and opera houses of the world. The perception of sound as music and the *psychological* dimensions of sound (*pitch*, *loudness*, *timbre*) will be further discussed in chapter 3, accompanying a description of the auditory system.

Sound as physical stimulus

The essence of sound is the propagation of energy into a medium, usually the air. This energy is transmitted as a pressure wave. The vibration of an object acts on the air molecules surrounding it, compressing those immediately in contact with the vibrating object. In turn these molecules compress those adjacent to them, and this pattern propagates through space until the wave eventually runs out of energy. A wave of energy is transmitted through the medium by successive *compression* and *expansion* (or *rarefaction*) of the molecules of the medium, be it water or air. Because both phases – compression and expansion – constitute departures from the 'resting' state of air pressure, the movements of air molecules begin to form an oscillation between the two phases. The result is a *sound wave*, a physical phenomenon that is capable of exciting the auditory mechanism so that a living organism perceives a sound.

Figure 2.1 shows how the vibrations of a tuning fork would lead to these kinds of disturbances. When the tuning fork is struck, its tines are set in motion. As a tine moves in an outward direction, the air molecules directly adjacent to it become compressed; as the tine moves back in past its midpoint, the molecules spread apart; and the process is repeated until the vibrations cease. With each oscillating movement of the tine, regions of oscillating high pressure (compression) and low pressure (expansion) are

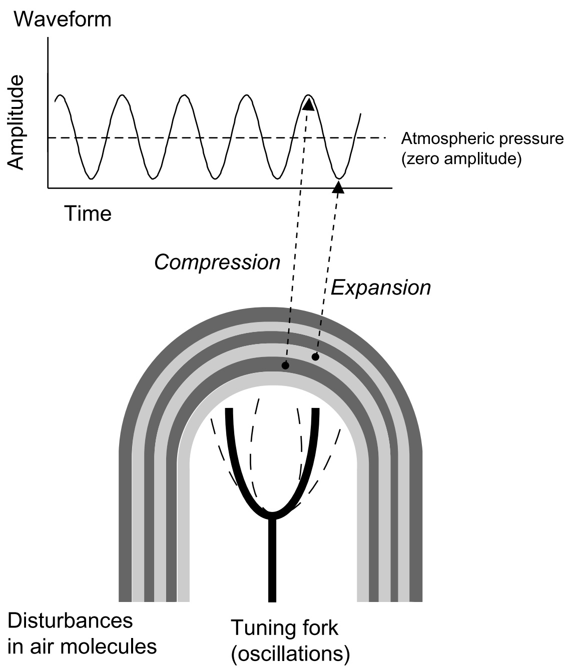


Figure 2.1 Propagation and representation of sound. Movements of the tines of a tuning fork (bottom) lead to alternating bands of compression and expansion among air molecules. For a pure tone (as in a tuning fork) the pattern of disturbances across time (for a given point in space) can be represented as a sine wave (shown at the top).

produced. In the figure, regions of compression are shown as dark bands to represent the clustering together of air molecules, and the spreading of molecules for regions of expansion is shown as light bands.

The motion of a sound wave is longitudinal. In a *longitudinal* wave, the movement of the oscillation is parallel to the direction of movement overall (as opposed to a *transverse* wave, characteristic of light, in which the oscillation is perpendicular to the overall motion). The movement of longitudinal waves can be demonstrated with a slinky or spring. When a momentary force is applied to the first coil, it nudges the coil adjacent to it, which in turn nudges the next coil, and so on. The disturbance travels through to the other end like a wave, but the coils themselves do not travel from end to end. Each coil makes an oscillating motion that contributes to the wave and seems to pass it along, and then returns back to its place. In a sound wave, molecular movement affects adjacent molecules so that a periodical pressure wave is propagated through the medium (i.e., air) – but the constituents of the medium (i.e., the molecules) move only locally.

Although the essence of sound is a disturbance, a single punctuated disturbance will not lead to the experience of music. Rather, the disturbances that lead to music follow regular patterns, resulting from *vibrations* in some sound-producing source. In turn, the air molecules begin to vibrate and the disturbance travels away from the source of sound, in an ever-widening wave of compressions and expansions, as shown in Figure 2.1. As it turns out, patterns of vibrations are integral to the link between sound and music. In the section that follows, we discuss different properties of these vibrations, and the sound waves (ultimately music) that result from them.

The properties of sine waves

A tuning fork produces vibrations of the simplest sort, resulting in a smooth *sine wave* (as shown at the top of Figure 2.1). Hardly any other sounds we hear in the real world can be fully described as sine waves or 'pure tones.' Nevertheless, a sine wave is a clear way of representing important characteristics of sounds. For instance, sine waves show clearly the alternating patterns of compression and expansion described earlier (see Figure 2.1). Two parameters of sine waves that are critical for our discussion of musical sound are illustrated in Figure 2.2.

One parameter has to do with the rate at which a sine wave alternates between compression and expansion: *Frequency* (*f*, the rate at which the crests or troughs of the wave pass a point in a given measure of time), which is measured in hertz (Hz) or *cycles* (of compression and expansion) *per second.* Thus a sound source that is vibrating at 50 cycles per second is said to have a frequency of 50 Hz. In Figure 2.2, sine waves plotted in the top row have a lower frequency (1 cycle per second) than those in the bottom row (2 cycles per second). Both signals would be below the range of human hearing.

In 1563 the mathematician-physicist Giovanni Battista Benedetti discovered

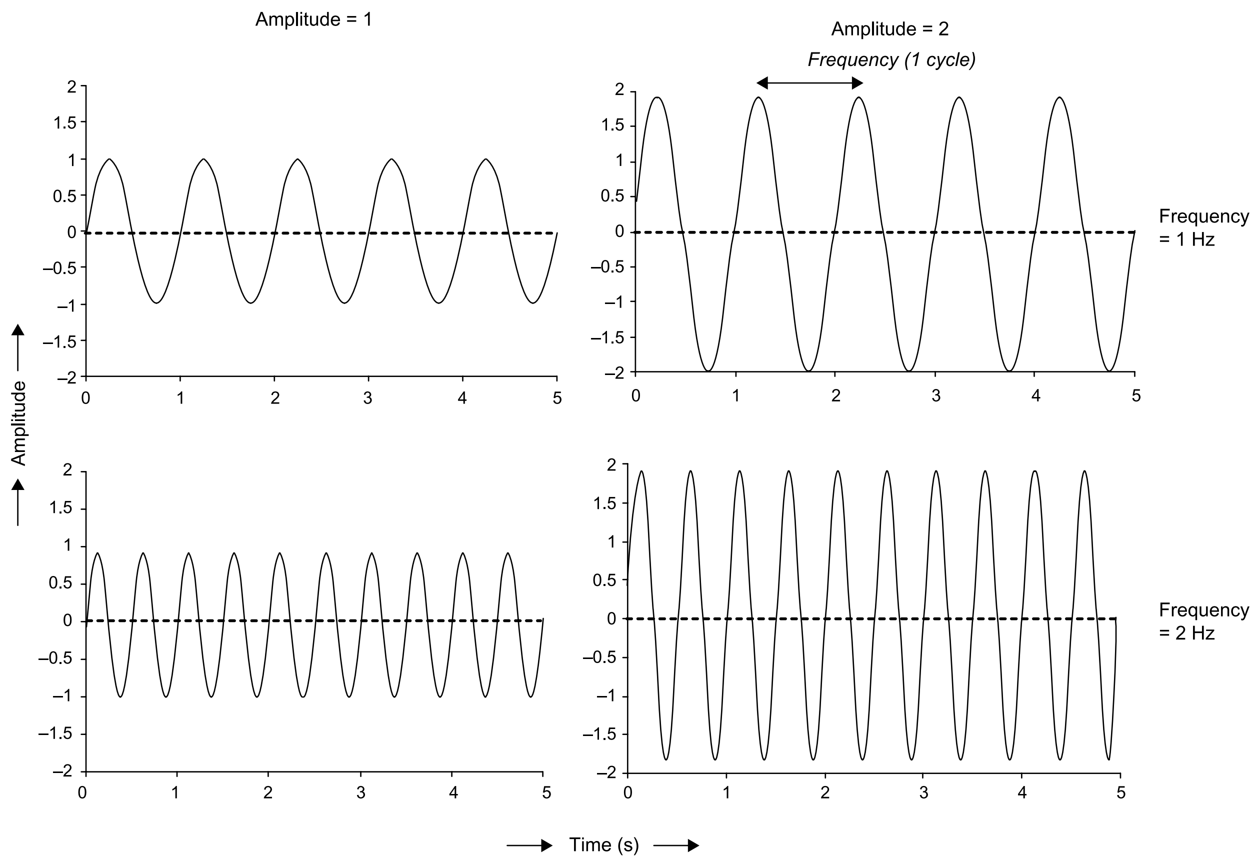


Figure 2.2 Four sine waves, varying in frequency and amplitude. Copyright © Peter Pfordresher.

that the *frequency of vibration* of the source of a sound gives rise to the aural sensation we call *pitch*. (Previously, musical theorists had followed Pythagoras in linking pitch to wave length.) When the frequency is high, there are many cycles in a given span of time, and the sound is perceived as high-pitched. A tone with a low frequency has few cycles in a given span of time and the sound is perceived as low-pitched. Extending the physical to the musical, consider the fact that changes in this simple parameter, frequency, can enable us to experience the opening phrases of 'Mary Had a Little Lamb' or 'Twinkle Twinkle Little Star' as distinctive melodies.

The second parameter of a sine wave that is relevant to a discussion of music is amplitude. The *amplitude* of a wave is the maximum displacement compared to the resting state, and corresponds to the psychological dimension of sound that we experience as *loudness*. Generally, the greater the amplitude of a wave, the more energy it transmits. Sound waves with more energy are generally perceived to be louder than sound waves with less energy, though the subjective effect depends to some extent on the frequency of the sound. In Figure 2.2, sine waves on the left side have lower amplitudes than those on the right side.

Complex sound waves

Sine waves, as shown in Figures 2.1 and 2.2, have only a single frequency of vibration, which is why they are often called 'pure' tones (just as light with only one wavelength is called a pure color). But, as we will see, most sounds are best understood as a combination of many frequencies that occur simultaneously. Such sounds are referred to as *complex tones*, and the interplay among their component frequencies leads to a quality of sound referred to as *timbre* – which helps us differentiate between the sounds of a flute and a clarinet, or to identify a voice on the telephone. The pure tone produced by tuning forks, beeps produced by automated teller machines and old computers are well described by sine waves, but a friend's voice, the sultry saxophone, and other sounds of nature require a more complex description.

Going back to the basic premise that sounds start with a pulse of energy in the form of vibrations, consider the objects that do the vibrating in musical instruments: strings tethered to wooden boards, pipes of different lengths, curved brass metal sheets, and so on. These are complex structures that, when stimulated, will not lead to a single simple frequency of vibration. Rather, the pattern of vibration will produce a multitude of frequencies that lead to a single perceived tone. Such tones are referred to as *complex*. The waveform for a complex tone is shown in Figure 2.3.

The appearance of the complex wave shown at the top of Figure 2.3 is quite different from the sine waves in Figures 2.1 and 2.2. However, any wave pattern, no matter how complex, can be decomposed into constituent sine waves. The process by which the relevant waveforms are abstracted from the complex pattern is referred to as *Fourier analysis*, named after French

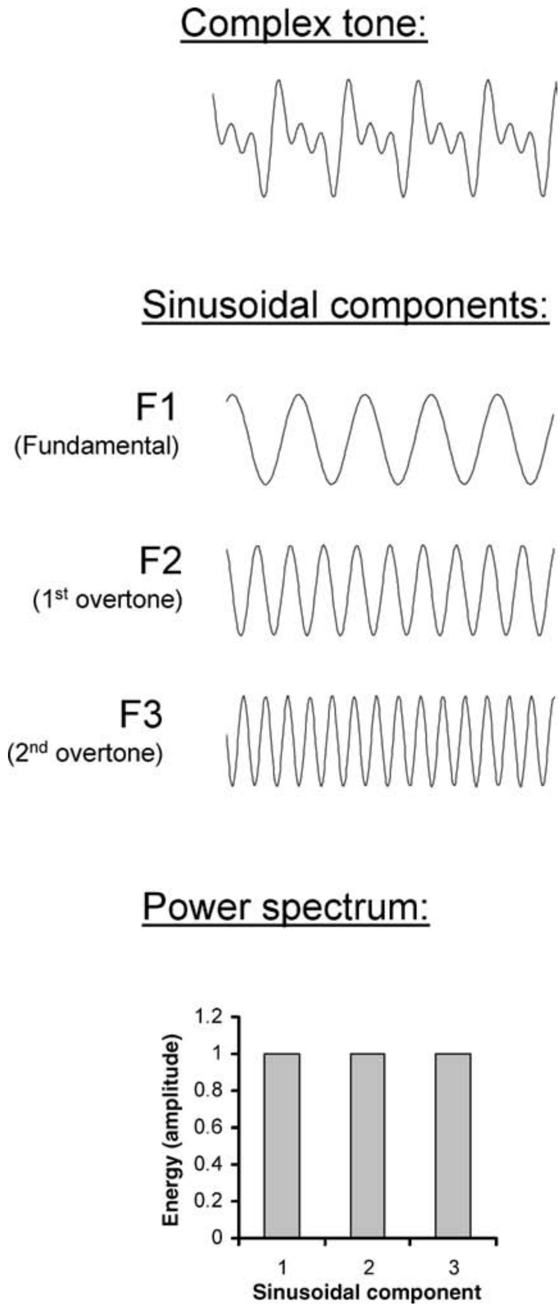


Figure 2.3 A complex wave, and below it a Fourier analysis. Copyright © Peter Pfordresher.

mathematician and physicist Jean Baptiste Joseph Fourier (1768–1830). Fourier showed that any function could be expanded into a series of sine functions.

The output of such a Fourier analysis is shown for the complex wave in Figure 2.3. Fourier analysis could be used here to represent the complex wave as the sum of three sinusoidal waves. Each of these individual 'components' is shown below the complex wave. Mathematically, the complex wave is simply the sum of each component. The component with the lowest frequency (longest wavelength) is commonly referred to as the *fundamental frequency*. Higher frequency components are often referred to as *harmonics* or *overtones*. Note that we have chosen as an example a tone that is not 'very' complex in comparison to a sine wave, having only three components. If one were to play a sound like the one shown in Figure 2.3 it would sound more like a buzz than a musical sound! Later we will demonstrate the complexity of 'real' complex tones.

The relationships among the frequencies of sine wave components often map onto simple, integer relationships. For the example shown in Figure 2.3, the first overtone is double the frequency of the fundamental and the second overtone is triple the frequency of the fundamental. Such is the case with a harp string, for instance. When it is plucked it does not just vibrate as a whole but also in halves, thirds, quarters, fifths and so on – each creating successively higher frequencies based on these orderly ratios (i.e., a series of harmonics or overtones). These relationships are often represented with respect to their corresponding musical pitches. For instance, the *harmonic* or *overtone series* based on the fundamental C2 (second highest C on a grand piano) is provided in Figure 2.4. The first 11 overtones are shown, but this series often

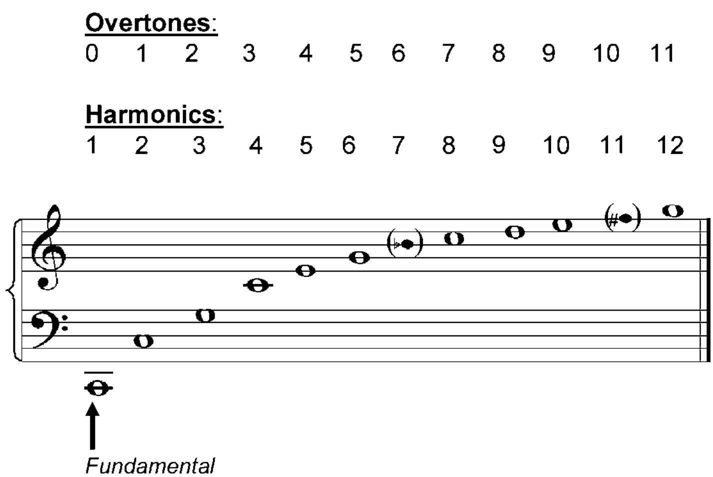


Figure 2.4 The harmonic or overtone series based on the fundamental C2, with harmonic and overtone numbers shown. Notes shown in parentheses denote approximate pitches.

continues beyond the perceptible range of human hearing. Note that two numbering conventions may be used. One convention is based on the *harmonic* series, for which the fundamental is the first (numbered F1) of several harmonics. The other convention assigns the fundamental the number zero (F0) and higher frequencies are numbered *overtones* starting with the first overtone. This scheme is more commonly used in the music literature.

Not all harmonics necessarily emerge with the same loudness. If a harp string is plucked in the center of the string, it is mainly the fundamental and the odd-numbered harmonics (even-numbered overtones) that will sound. Clarinet tones in the low register are almost completely lacking the second harmonic. The sound each instrument produces is made up of a different composition of frequencies that in part give it a characteristic sound. This is another way in which Fourier analysis contributes to our understanding of sound. The bottom section of Figure 2.3 shows a representation of sound referred to as a *power spectrum*, which plots the amount of energy (or power) associated with each harmonic in a tone (expressed as amplitude). Because energy is related to amplitude, the power spectrum can be thought of as a way of showing the amplitude of sine waves in a complex tone. The artificial tone in Figure 2.3 consists of components that are equally ‘powerful.’ Later we will see examples of natural tones for which each component is not similarly powerful.

Sounds like the one shown in Figure 2.3, or a plucked harp string, are often called *harmonic* complex tones because the overtones are integer multiples of the fundamental frequency. Such integer-based relationships are characteristic of sounds with a pitch-like or ‘melodious’ timbre. On the other hand, sounds with a rough or ‘noisy’ timbre have overtones that are not so simply related to the fundamental. Figure 2.5 shows an example of a complex tone (represented as in Figure 2.3) made up of sine waves that are not based on integer relationships. If played, the tone in Figure 2.5 would probably sound ‘rough.’ For instance, the static noise of a radio that is not tuned to a station comes from frequencies that are entirely unrelated to each other. In the human voice, a smoother timbre (such as that of crooner Michael Bublé) is closer in structure to the kind of sound wave shown in Figure 2.3 whereas a rougher sounding singing voice (perhaps Louis Armstrong or Tom Waits) would have more components that do not conform to such integer relationships. Likewise, many percussion instruments (like drums and bells) include harmonics that do not conform to integer relationships.

Thus far we have emphasized acoustical differences *across* instruments (and singers). However, it is important to note that the spectral structure of a single instrument is not constant but can change as a function of pitch, loudness, duration, or expressive intentions. An example of how pitch influences the power spectrum is shown in Figure 2.6. This figure was generated by recording a trombonist producing B♭ in three different octaves. The left side of the graph shows (complex) waveforms associated with each pitch and the right side shows their power spectra (cf. Figures 2.3 and 2.5). Each B♭ is

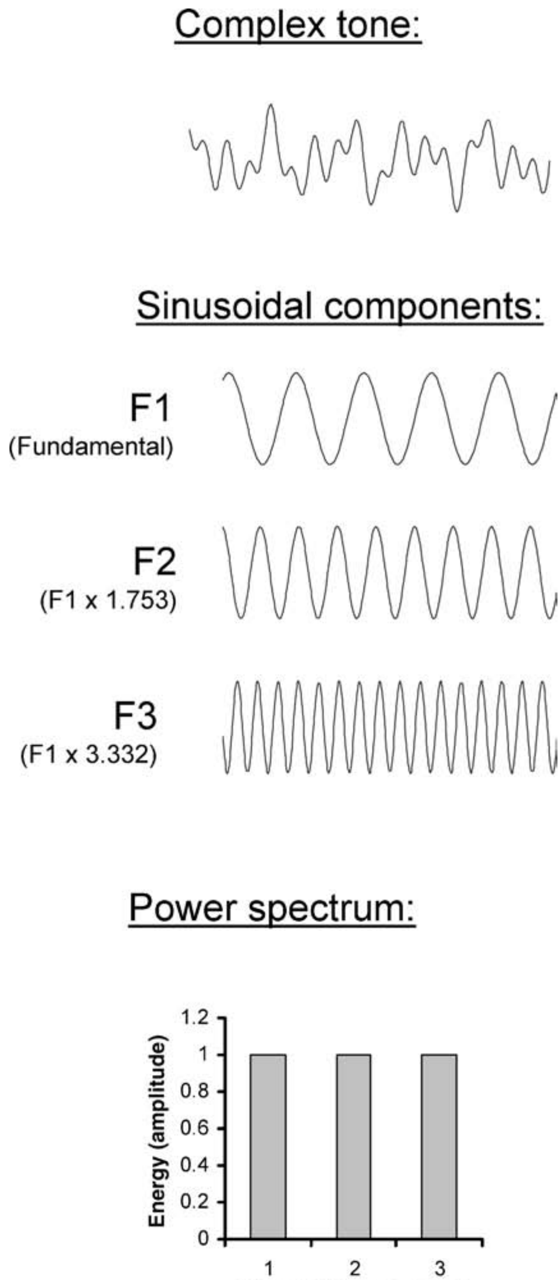


Figure 2.5 Waveform for an anharmonic complex tone, with its Fourier analysis below. Copyright © Peter Pfordresher.

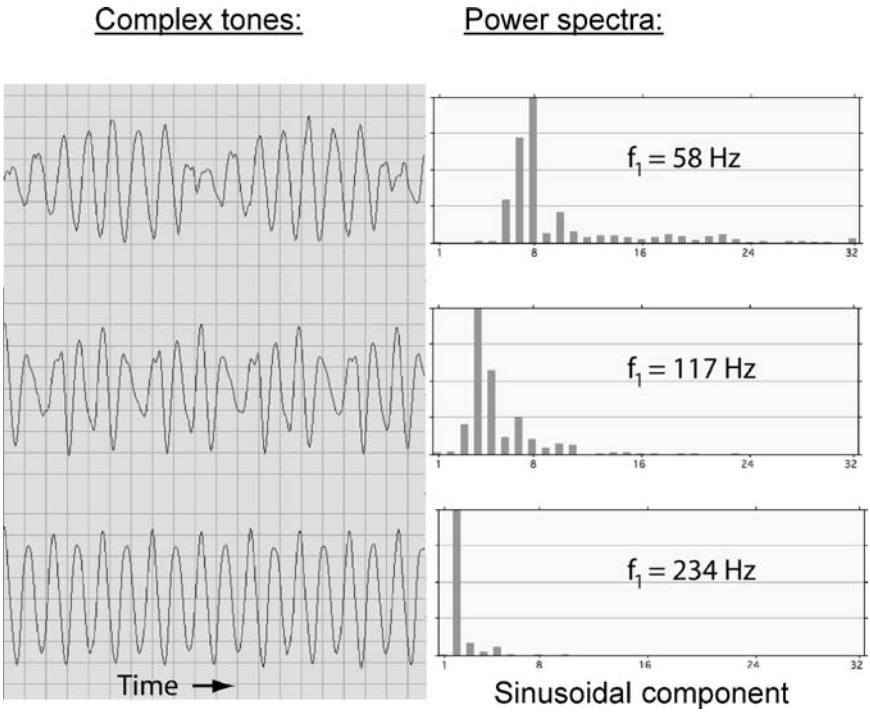


Figure 2.6 Waveforms (left) and power spectrum plots (right) for three tones produced by a trombone. (Power spectrum plots show frequency values as harmonic number rather than Hz.)

Source: Figure created by Matt Segars and composited by Rod Nave at the Department of Physics and Astronomy at Georgia State University, used with permission.

shown as F1 in the power spectra, along with its upper harmonics (or overtones). In principle, the power spectrum for each produced pitch could be identical because frequencies are plotted by harmonic number (1 for fundamental). However, this is not the case. The acoustic structure emphasizes relatively lower frequencies within the range of frequencies presented when the performer produces a higher pitch, and emphasizes relatively higher frequencies when the performer generates a lower pitch. These differences ultimately reflect the fact that a certain (intermediate) range of frequencies is likely to have a higher level of energy across many different produced pitches in the trombone. This range ends up being relatively high when one plays a low pitch and relatively low when one plays a high pitch. Such bands of high energy are known as *resonances* and may help the listener identify a specific instrument. We discuss the concept of resonance more in the next section.

Fourier analysis involves complex mathematics for which we have given a verbal sketch. Readers interested in this topic may wish to refer to Benson's

(2006) *Music: A Mathematical Offering*, which includes a good chapter on Fourier analysis in a musical context, or Kammler's (2008) *A First Course in Fourier Analysis* (which, despite its title, is an in-depth exploration with applications to music and other areas).

The acoustics of musical instruments

Musical instruments (and we include the human voice among them) are devices for producing sound waves. Each species of instrument has its own acoustic characteristics, and the physical structures of instruments vary widely. However, almost all musical instruments depend on the principle of *coupled acoustics*. That is, in most musical instruments there are two vibrating devices: One generates the sound wave and the other is *coupled* to it, and amplifies it.

The initial vibration that is activated by the performer – for instance by blowing into an oboe or bowing a violin – is often of insufficient energy to be properly heard in a large space or at a distance. A *resonator* (such as a hollow tube, an air chamber, or a flexible soundboard) is 'coupled' to the basic device. The importance of coupling can easily be demonstrated by blowing the reed of a clarinet mouth-piece detached from the instrument, which will only produce a thin squeak. In the clarinet, the vibrations of the reed excite oscillations in the column of air in the body of the instrument. The air column is the resonator. In the double bass, the vibrations of the strings resonate in the hollow body of the instrument. The similarity between the terms *resonator* and *resonance* (introduced earlier) is no coincidence. Any given resonator is likely to serve as a better conduit for certain frequencies than others. These resonances, as mentioned before, help give a particular instrument or voice a characteristic timbre across many different pitches and loudness levels.

The groundwork for what we know about sound resonators and many other acoustic phenomena was laid by Hermann Helmholtz in 1863, with his publication of *On the Sensations of Tone*. One of the leading physicists of the nineteenth century, Helmholtz's interest in music led him to propose a research paradigm for bringing the physical sciences into a working relationship with the arts. Realizing that the discoveries of physical acoustics (such as the ratios of modes of vibration of a taut string) were correlated with experiences of consonance, harmony, and other phenomena that were the basis of music, he looked for a bridging science. He found this in physiological acoustics, the anatomy and physiology of the ear, and neural links to the brain.

Most pertinent to the present discussion is Helmholtz's experimentation with sound resonators consisting of a large collection of hollow spheres (of many different sizes) made of glass or metal, with two tiny openings on each end. One opening was to be placed snugly inside the ear for an air-tight fit, and the other was to be pointed at the source of a sound. Helmholtz discovered that if the sound contained a harmonic with the same or similar frequency as the natural resonant frequency of the hollow space inside the resonator, the harmonic would be amplified and would sound inside the resonator. Just as *Helmholtz resonators* of different sizes and materials amplify different harmonics, the properties of the resonators of musical instruments in part determine which harmonics are amplified or suppressed. 'Brighter' tones occur when the higher harmonics are present and amplified, and a 'duller' sound is produced when upper harmonics are weak or missing.

Strings

We can group instruments in which sound is produced by setting a string into motion into three main groups. In one group there are 'bowed strings.' These are the instruments such as violins and double basses in which a string is continuously supplied with energy from a moving bow, and the vibration is maintained at the same pitch to give an enduring tone. In another group such as harpsichords, guitars, and mandolins, a string is set in vibration after brief contact with a plectrum or some other device that plucks the string (in the case of Jimi Hendrix, sometimes his teeth!). Then there is the third group of instruments in which the string is struck with a hammer, as in the piano which also belongs to the family of percussion instruments. In most of these instruments, the hammer rebounds almost immediately off the string. Whether or not the string continues to vibrate depends on whether it is damped by contact with some material such as felt.

Not all vibrations of a string are controlled by the actions of the performer. One of the three strings of the long-necked string instrument, the Indian *sitar*, is not plucked but vibrates in sympathy with the vibrations of the melody and drone strings. A similar though subtler effect of *sympathetic vibrations* can also be demonstrated on a grand piano when playing one or more keys while pressing the sustain pedal (the rightmost pedal). The pedaling action removes the dampers from the strings, freeing them to vibrate in sympathy with the strings corresponding to the depressed keys.

Despite the variety of ways the strings are set in motion, the sound we hear as music comes from the same principle of coupled acoustics. The vibrating strings radiate a weak sound initially, but the energy is transferred to the bridge and top plate, and into the body of the instrument and the back plate. In addition to other factors such as the thickness, length, mass and tension of the strings, the particular characteristics of the resonating body play a critical role in the resulting sound's timbre.

Woodwinds and brass

The same coupling principle that lies behind the production of musical tones by stringed instruments also applies to wind instruments. A vibration produced in the mouth-piece sets a column of air resonating. It is this vibration that is transmitted to the listener by the air. 'Edge tones' are created by an action such as blowing across a hole and onto a sharp edge. The tones are produced because the upper and the lower planes of an edge produce asymmetrical vortices as the air stream passes over the edge. This asymmetry induces an oscillation in pressure, which is picked up and amplified by the adjacent air column in the body of the instrument. The effect only works if the incoming air stream is narrow, as for example from the pursed lips of the flute player (essentially a hole), or through the embouchure of the clarinetist (which is essentially a slit). This is another example of coupled acoustics because the physical process at the edge is quite different acoustically from the wave transmission in the air column. But the instrument only produces musical tones because the two processes are coupled.

'Reed tones' are produced by the player setting one or two reeds vibrating, which in turn produces sound waves in the adjacent air columns. This is the system to be found in single-reed instruments such as clarinets and saxophones, and double-reed instruments such as oboes and bassoons. In the Scottish bagpipe, the player does not put the reed into the mouth, but blows into a reservoir in the form of a bag; when the bag is squeezed, sound is created as air passes through the reeds of three pipes creating drones and a fourth 'chanter' pipe providing the melody. Brass instruments use 'reed tones' but the 'reeds' are the human lips; the vibrating motion that creates the sound is independent of the instrument, being freely controlled by muscle tension. Brass instrumentalists must expend a lot of effort developing the muscles of their lips as a way of controlling pitch height and intonation (good brass instrumentalists can play music simply by 'buzzing' their lips). Finally, we should notice that the human voice is a reed tone instrument, the vibrations of the vocal cords resonating in the air chambers of the respiratory system. The acoustics of the vocal resonators are complex; imagine the constantly changing shape of the resonating chambers of a vocalist singing a lyric such as 'Hallelujah!'

There are interesting variations in tone quality depending on whether the body of a wind instrument is of conical bore or even bore throughout, and how it flares at the end. A flared bell radiates the high harmonics, which in part gives the trumpet its 'bright' sound. The length of the resonating columns of wind instruments also varies widely; generally, longer pipes produce lower or deeper tones. The tubing of a French horn when unrolled into a straight pipe, for instance, measures about 2.7 to 3.7 meters (depending on the type of horn). This is approximately two to three times the length of the pipe of a trumpet, accounting in part for the warm, rich tones it can produce. It is the length, shape, and materials of the pipe resonator that, to a large extent, account for the distinct sounds produced by different wind instruments.

Percussion

Unlike the violin or the clarinet, in which the performer continuously supplies energy to the system, percussion instruments receive energy in short bursts. A drum stick strikes the membrane of the drum in a single stroke. This stroke sets the relevant parts of the instrument vibrating at their natural frequencies, depending in part on the size and rigidity of the materials with which it is made. As the surface of the drum is flexible, the impact also sets into motion a wave disturbance that travels to other parts of the surface that were not directly struck. Thus a single strike of a drum generates many frequencies that bear no simple relationship to each other, producing an 'unpitched' sound. The induced vibrations then die away until the performer makes another stroke.

Drums, chimes, and marimbas, and many percussion instruments have resonating chambers or tubes. The coupled acoustics of the *saron*, one of the pitched percussion instruments in the Indonesian *gamelan* orchestra (described in chapter 15), can easily be deconstructed as it consists of a few loose metal keys, simply placed over a shallow wooden box, which serves as a trough resonator. Without it, the saron would barely be heard over the full-percussion *gamelan* that often performs in outdoor venues. The piano's main resonator is the soundboard; essentially, it repeats the vibratory motions of the piano strings and must be carefully crafted if the instrument is to produce a rich sound. However, not all percussion instruments have resonators. For instance, vibrating plates such as cymbals and gongs have no built-in 'amplifiers.'

Across the huge diversity of musical instruments which human ingenuity has created, the basic principles are the same. With very few exceptions, instruments produce pitched sound waves, from which the listener extracts the basic musical qualities of pitch, loudness, timbre, and variations in duration (rhythm). Almost all instruments depend on the principle of coupled acoustics to generate sound waves that are clearly audible at some distance from the performer, and the various resonators contribute their characteristic properties to the sound wave that ultimately emerges. However, the quality of sounds produced by the instruments and voices is also affected by the characteristics of the physical environments into which they are released. Next, we consider the sound fields in which music is commonly performed and appreciated: concert halls, opera houses, and other performance venues.

The acoustics of musical venues

In 1895, Wallace Clement Sabine, professor of mathematics and philosophy at Harvard University, was asked to solve a practical problem: It was hard to understand what speakers were saying in a lecture hall in Harvard's Fogg Art Museum. After performing many tests, Sabine determined that sounds in the hall were sustained within an audible range for a long duration before decaying, making speech indistinct. Sound-absorbent materials were placed in the room and the problem was remedied. Prior to this time, buildings were not usually constructed with acoustics in mind. The first music hall to be designed with the help of acoustic engineering was the New Boston Hall (Boston Symphony Hall) in 1900, for which Sabine applied formulas he developed for calculating ideal reverberation times. He was soon regularly consulted to assist in building design and address acoustic problems of completed structures. Sabine's work laid the foundations for *architectural acoustics*.

Today, the study of the acoustics of a concert hall or opera house has come to encompass more than just the (rather complex) physics of sound propagation in an enclosed space. It has come to include the consideration of features that affect the perception of music and often also the subjective evaluations of the listeners. The best conditions for hearing an orator can be fairly well defined in acoustic terms, as the focus is often on the clarity and intelligibility of speech. However, a host of other factors, such as the type of music being performed and even the reputation of the music hall, are relevant to how music sounds to a particular listener or performer. Following important work by Yoichi Ando (1985) and others on the *subjectivity* of architectural acoustics, much of the research on the acoustics of music performance venues takes listeners' and performers' subjective preferences into account.

Studies of the quality of experience of music under different conditions can be carried out in many different ways. It is possible to create artificial sound fields with properties that mimic the structures of the sound fields one would find in a concert hall, church or other enclosed space. For instance, different materials may be introduced into an anechoic chamber to shape the way sound fields are created, and listeners may be asked to judge the quality of the sound. Another method is to create binaural musical recordings (employing a model of a human head with stereo microphones inserted in each ear, in order to record sound traveling to both ears) to quite faithfully reproduce musical performances in different halls. Listeners can then compare these recordings inside an anechoic chamber. Neither of these procedures, however, captures the complete experience of sitting in the performance hall during a live concert!

An alternate approach is to ask listeners to give their impressions of the acoustical qualities of music performance halls of which the sound fields and acoustic parameters have been studied. Leo L. Beranek (2004) used this approach in an extensive study of 50 of the major concert halls and opera houses of the world, originally reported in 1962 and subsequently revised in an expanded volume in 2004. By developing a kind of composite vocabulary of musico-acoustical terms, he set about interviewing conductors, music critics, and experienced listeners on their experience of music in various venues. Then, using standard acoustical concepts and techniques, he also measured the physical attributes of the halls. As we shall see later in this chapter, he found some consistency in listeners' and performers' preferences in some of the world's great music halls.

Direct and reflected sound

Our discussion of the acoustics of venues for musical performance focuses on a few basic concepts as they apply to the quality of listeners' and performers' experience in music performance venues: directed and reflected sound, sound absorption, and reverberation time. Both physical (or objective) and subjective parameters are considered.

Direct sound travels directly from the source to the listener; it contains auditory information in an uncontaminated form. The clarity of the sound of an orchestra depends heavily on direct sound. *Reflected sound* reaches the listener by 'bouncing off' one or more surfaces such as walls, ceilings, pillars, and sound baffles. Depending on the time interval between the arrival of direct and reflected sound at the ears of the listener, the added reflected sound may add a pleasant richness to the musical tones – or it may 'muddy' up the sound and even create distracting echoes at only 50 millisecond (50/1000ths of a second) delays at high reflection levels (Gade, 2007). The *initial-time-delay gap* (*ITDG*) is defined as 'the time at which the first reflection is heard after the direct sound' (Beranek, 2007, p. 4). An ITDG of about 15 milliseconds seems to be optimal for most concert halls. If greater than about 35 ms, 'the hall will sound like an arena, with a lack of intimacy' (Beranek, 2007, p. 4).

There is also the clarity of sound to consider, or the extent to which musical tones or a singer's lyrics sound clear and distinct. The *clarity index* is a measure of the ratio of early sound energy (from *early reflections*, which arrive within the first 80 milliseconds of the direct sound) to late sound energy (*late reflections* arriving after 80 ms). The preferred clarity index of a performance venue varies with the characteristics of the music being played; contrapuntal Baroque music demands higher clarity than music of the romantic period (Reichardt, Alim, & Schmidt, 1974, p. 243).

Two-thirds of the world's concert halls that were most highly rated by conductors, music critics, and experienced concert-goers in Beranek's (2004) study are 'shoebox'-shaped. The narrow width of the halls allows for stronger *lateral (side-to-side) reflections* to complement the sounds coming from directly in front of the listener (from the stage). This creates the sensation of being 'bathed' in the sound, which is preferred by most listeners (Ando, 1985). This is not to imply a long narrow hall with flat walls and plain ceilings. Sound is dispersed more uniformly throughout the space when there are some irregularities in the 'box,' such as coffered (ornamental, recessed) ceilings, balconies, statues, and textured surfaces that can diffuse or diffract sound waves.

Fan-shaped halls are often more problematic. The fan shape leads to progressively widening walls toward the back of the hall, which directs lateral reflections away from the listeners. This weakens the sense of *acoustic intimacy* (the degree to which sounds seem to be coming from nearby rather than remote surfaces), *listener envelopment* (the sense of being surrounded by reverberant sound), and *warmth* (provided by transmission of low frequencies such as bass tones). Further, while the first reflections arrive from the narrow parallel walls in shoebox-shaped halls, the first reflections from halls that fan out from the stage are likely to come from the more distant path of the tall ceilings – thus making for a longer reverberation time. Concave walls behind the audience also tend to send reflected waves back to a point on the stage and can create echoes, an annoyance to many performers (especially instrumentalists).

To give an interesting case study, a notable exception is the fan-shaped *Aula Magna* at the University of Caracas in Venezuela, which receives good reviews from performers and critics. To address the potential acoustical problems stemming from the nonparallel walls and curved features of the architectural design, it was decided during construction that sound-reflecting panels often referred to as 'clouds' (covering an area equivalent to 70 percent of the ceiling) were to be hung below the ceiling and along the side walls. Rather than using standard rectangular panels, sculptor Alexander Calder was commissioned to create suspended panels in the dynamic abstract shapes characteristic of the geometric-shaped mobiles for which he is known. The result, shown in Figure 2.7, is both acoustically and aesthetically pleasing!

However, pure laws of physics alone do not always predict audience preferences. Eight of the nine top-rated opera houses in Beranek's (2004) study are horseshoe-shaped. He noted that the horseshoe is not acoustically ideal (the curved wall behind the audience directs sound back to the center of the stage, and balconies create areas of low resonance underneath them), and

Figure 2.7 The interior of the Aula Magna at the University of Caracas in Venezuela, featuring 'clouds' by sculptor Alexander Calder.

Source: Published courtesy of Leo Beranek, *Concert Halls and Opera Houses* (Springer, 2004).

yet seemed to be preferred by listeners and performers for other qualities. For example, this arrangement brings performers and audience closer together, which increases *visual clarity* so that facial expressions and gestures can be clearly seen, and *acoustical clarity* for the intelligibility of the lyrics and speech. The slight echo common in halls of this shape provided feedback to singers, and did not seem to distract listeners.

Sound absorption

Not all sound waves reach the ear directly (direct sound) or bounce back from dense surfaces (reflected sound) in a contained space. Some become trapped in materials such as ceilings and walls, stage curtains, carpeting, and seat upholstery – and reflect little energy back. This sound has been *absorbed*.

Absorption materials commonly found in music performance venues can be classified into two kinds. *Porous absorbers* (curtains, theater seats, carpets) absorb high-frequency sounds more efficiently than bass sounds. *Resonant absorbers* (such as wood panels) are set into vibration by the energy released by the sound source and respond to low-frequency sounds. By conducting numerous tests of the sound absorption of different materials, Sabine determined that *absorbing power*, the key to reverberation times, varies not only with the material reflecting sound but with the frequency (pitch) of the instrument. For example, the absorbing power of a given surface for the higher register of the violin is nearly double that of the same surface for the lower register of the double bass (Sabine, 1922).

Audiences also absorb sound; imagine the heavy fabrics of formally-attired concert audiences of the 1700s and 1800s! Audience absorption is a tricky variable as one cannot always predict the number of occupants that will attend a performance, nor the exact arrangement of occupants filling the available seats. Audience area is larger if the audience is seated on a slope, and thus absorption is affected not only by the size of the crowd, but by the dimensions and plane of the seating space. The seats themselves also absorb a lot of sound, which is one reason why back rests in most concert hall seats are usually low (ending below the shoulder) and chairs are not nearly as plush as seats in a movie theater.

Reverberation time

Sabine was a pioneer in identifying reverberation time as a critical factor in indoor acoustics. Sabine's Law (the formula for calculating reverberation time, T) is written as:

$$T = 0.161\, V/A$$

where V is the volume of the room in cubic meters, and A is the total absorbing power in sabins. Sabine's measure of reverberation time was the time it takes a sound to decay until it is barely audible, as sensitive methods to determine sound levels had not yet been devised. He made painstaking systematic measurements of the time it took sound to decay, starting from 1,000,0000 times the first audible sound level to silence (Sabine, 1922, pp. 60–68). Today, *reverberation time (or T)* is defined as the length of time for a sound to decay by 60 decibels. Although alternate measures such as EDT (early decay time, based only on the initial part of the decay) have since been devised, *T* is still regarded as one of the most important basic acoustic parameters (Gade, 2007).

Performance halls range from reverberant to dry – that is, from those in which there are many complex reflections and those with absorbent walls and ceilings, which reduce the loudness and longevity of these reflections. From the perspective of performers, two important concerns are *ease of ensemble* (the degree to which performers can hear and play together) and *support* (the degree to which the room facilitates the musicians' efforts to create tones and fill up the space). Early sound reflections in the stage area are critical for ease of ensemble, while both early and late reflections are important for good support. When performing in a 'dry' space, it is difficult for instrumentalists and vocalists to feel like they are 'filling up' the room with music, as the sound disappears quickly. Instrumental and vocal ensembles must also give more precise performances in dry rooms, as asynchronies are more discernible. A choir singing lyrics such as 'to tell a tale of tragedy' must really stay together to sound crisp in a dry space! On the other hand, long reverberation times in large spaces may make it difficult for performers to hear themselves and each other, as the *horizontal* and *vertical clarity* (the distinctiveness of tones played successively and tones played simultaneously) are diminished.

The highest rated concert halls in Beranek's (2004, 2007) study have a reverberation time of 1.8 to 2.0 seconds, while a shorter *T* of 1.24 to 1.6 seconds seems to be ideal for the best opera houses. Venues designed for speech (such as oration, and plays) generally require the shortest *T*s of 0.7 to <1.0 second, as it is critical for the audience to perceive fine differences in speech sounds. In listening to music, however, the *loss* of some sounds (the rasping of the cello bow or the hiss of air expelled from a clarinet) due to longer reverberation times may actually enhance the listener's appreciation of the music. Some degree of blending of the sounds of an orchestra or choir is essential to a music performance. Auditoriums with modifiable size and flexible reverberation times (e.g., with chambers that can be opened or sealed, or removable panels) are practical designs for performance venues that must accommodate speech, as well as vocal and instrumental performances.

Optimal reverberation times vary for different instruments and vocal ranges, and musical repertoire. Even within the same 'family' of instruments, ideal reverberation times differ somewhat; pianists prefer halls with shorter reverberation times whereas pipe organists prefer more reverberant spaces (Veneklasen, 1975). *La Scala* in Milan has a *T* of 1.2 seconds, while Wagner's *Festspielhaus* in Bayreuth has a much longer *T* of 2.2 seconds. It is not hard to suggest reasons why this difference should emerge, considering the differences between Italian and German opera! One might compare the lucidity of Verdi's operas with the lush, expansive orchestration of Wagner's operatic works.

Very large auditoriums are rarely effective for music performance. In his study of architectural structures for the performance of music from the seventeenth century to the present, Forsythe (1985) notes that the Royal Albert Hall in London (built in 1871) was *ten times* larger in volume than most concert halls of its time, and attributes many acoustical problems it encountered through the years to its immense proportions. In very large modern spaces such as indoor arenas which may house audiences as large as 10,000 or 15,000 or more for 'pop' and rock concerts, performers must rely on electronic amplifiers and loudspeakers to carry and disperse sound, and low-frequency reverberation is particularly difficult to control. Generally, low-frequency absorbers are not as efficient as those that absorb high frequencies, which is why audiences in such venues often find themselves in a wash of persistent low-frequency sounds.

Our discussion has focused on how considerations about music performance have shaped the way we design and construct buildings, but the effects may have been bidirectional. Sabine once argued that the acoustics of a sound space are so critical to effective performances that the architectural traditions of different people and different eras may have fundamentally shaped the development of music (cited in Forsythe, 1985). The cavernous, highly resonant stone buildings of the Romanesque period allowed vocal tones to linger, supporting the exploration of rich vocal harmonies characteristic of choral music of that time. As the classical outdoor amphitheater evolved into the roofed horseshoe-shaped concert building, the improved horizontal clarity may have lent itself to the development of ornate contrapuntal Baroque music with its complex interplay of melodies (Forsythe). It is intriguing to consider the possible *mutual* influence of the construction of musical buildings and the construction of musical works.

Coda

At first glance, the science of acoustics does not seem very daunting. Wave trains in the air and the harmonics of vibrating strings are not difficult concepts to master. Acousticians have succeeded in extracting many listener-relevant properties of such sound fields, so that we can form fairly clear ideas of how musical experience is related to the physical properties of the field. The complexity of sound fields is directly related to the structure and properties of the bounded spaces in which music is played and heard. But here, mathematical analysis and ascertainable principles begin to part company. Using the broadly defined acoustical concepts defined in this chapter, it is possible to draw some correspondences between architecture and audience. However, one of the lessons to be learned from the research on the acoustics of performance venues is how difficult it has turned out to be to design a ‘perfect’ acoustic environment from the principles of the science of acoustics! There is much about the experience of performing and listening to music that cannot be captured by physical laws alone. It is to the psychological and more subjective qualities of musical sound, and the pathway from the source of a sound to the ear and auditory cortex, that we turn in our next chapter.

3 Sound and the neurophysiology of hearing

Passport photographs for many countries require both ears to be clearly visible. This is because the ears, though not quite as unique as fingerprints or snowflakes, are distinctive features of a person's appearance. If you were to gather a dozen of your friends and peer closely at their ears, you would be likely to find great variability in the shape, size, and particularly in the convolutions or patterns of ridges, of their ears! As we shall see, the distinctive shape and irregularities of form of the visible portion of the ear serve important functions, as the outer ears gather sound waves and funnel them through the auditory canal to the fine structures of the middle and inner ear.

This chapter focuses on the workings of the marvelous ear, to reveal the mechanisms in the auditory system that enable people to hear the musically salient aspects of sound. We begin with an examination of pitch, loudness, duration, and timbre as the four dominant sonic properties of the heard sounds from which music emerges as an auditory experience. Some of these concepts were alluded to in the last chapter in terms of the physical characteristics of sound waves. In this chapter, we focus on their subjective consequences for the listener, as mediated by the ear and the auditory cortex. How do the ears and the relevant parts of the nervous system extract these features of sound for a person to hear?

Characteristics of single tones

The previous chapter described how a sound wave is propagated, and explained that sound waves can be characterized by their frequency, amplitude, and complexity. As discussed briefly in the previous chapter, these physical properties of a sound wave correspond to the psychological qualities of sound that we refer to as *pitch*, *loudness*, and *timbre*, as shown in Table 3.1. In addition, there is also the *duration* of the acoustic or auditory event, as it must unfold in time.

Table 3.1 Physical and psychological dimensions of a tone.

Physical properties of sound waves	*Psychological properties of tone*
Frequency (hertz)	Pitch
Amplitude (decibels)	Loudness
Complexity	Timbre

Pitch

The perceptual experience of pitch is related to the frequency of vibrations in sounds. The ability to perceive pitch thus relies on the ear's ability to encode frequencies from physical stimuli. The pitches that most human beings can detect range from about 20 to 20,000 *hertz* (or *Hz*, which refers to *the frequency of a wave expressed in cycles or oscillations per second*). This capacity varies widely from person to person and declines with age. After age 30 or 40, *presbycusis* (i.e., hearing loss associated with aging) results in a loss of sensitivity in the highest ranges. This type of hearing loss usually does not appreciably affect music listening, except perhaps in the quality of tonal richness provided by the highest harmonics, as most music is limited to the range of approximately *20 to 4000 Hz*. To give some idea of what this range covers we can think of a grand piano, which spans approximately 27 to 4200 Hz. At the extreme ends of the spectrum of musical tones, the lowest pipe in a pipe organ is about 16 Hz (and is felt more as a rumble or vibration than a tone) and the highest note of the piccolo may reach 4500 Hz. However, the threshold of human sensitivity to pitch extends far beyond the piccolo's top register; to consider the highest pitch that a college student with intact hearing can detect (20,000 Hz), imagine a grand piano extended by a little over two more octaves of keys.

In the previous chapter, we discussed how most sounds are complex, being comprised of multiple vibrating frequencies. Even a single musical tone, created by depressing one key on a piano, consists of many different frequencies. Yet the perceived quality is of one pitch. Indeed, one of the mysteries of auditory perception, directly relevant to music, is that we typically hear one sound as having one pitch, but can still make out distinct pitches when multiple sounds form a cluster (as when a pianist plays a chord). Much research has been devoted to understanding how the auditory system extracts a single pitch from a complex combination of frequencies – the reverse of Fourier analysis – leading to theories of pitch processing in the auditory system. Such theories have largely been motivated by the physiology of the cochlea. We will therefore discuss them later in this chapter, accompanying a description of how the cochlea transduces the sound wave.

The simplest account of pitch processing, which works in many but not all cases, is to say that perceived pitch is linked to the lowest (fundamental) frequency in a complex tone, while the upper harmonics contribute to the timbre of that tone (see chapter 2). However, this generalization does not always apply. Consider listening to someone singing a song on the telephone. The frequency range of most telephones is limited and as such does not go low enough for many fundamental frequencies of low voices. Yet a listener hears about the same pitch (though not the same timbre) as if the singer were physically present. This experience is connected with a curious illusion, the phenomenon of '*residue pitch*' or '*virtual pitch*' (Schouten, 1938; Terhardt, 1974). That is, under some conditions, a missing fundamental can be heard solely on the basis of hearing its harmonics. The same phenomenon is at play when listening to a pocket-size transistor radio. The fundamental frequencies of the lower tones are often not audible and yet the proper pitches are usually heard as the fundamental is 'inferred' from the harmonics.

Another important concept related to pitch perception, motivated by physiology but grounded in behavioral studies, is the *critical bandwidth* (or *band*). The critical band refers to a range of frequencies that evoke a similar response in the auditory system, and bears on the degree to which our auditory system responds selectively to different frequencies. Frequency is a continuous variable, with an infinite number of possible values, and so it would be inefficient for our auditory system to respond selectively to every possible value. So instead it seems that our auditory system responds similarly (e.g., with a similar cochlear response) to frequencies in close proximity to each other – frequencies that evoke a similar response are said to fit within the same critical band. In some cases, this 'shortcut' of the auditory system can influence our perception of music. The most prominent example is in the case of musical dissonance.

The terms *consonance* and *dissonance* refer to a basic and important subjective continuum in music – reflecting the degree to which tones that sound simultaneously (called a *harmonic interval*) sound pleasing or displeasing (Helmholtz, 1863). One paradox regarding consonance in music is that the fundamental frequencies of pitches in highly dissonant intervals – for instance the fundamental frequencies of pitches forming a tritone (e.g., C and F♯) – typically do not fall into the same critical band. The resolution to this paradox lies in the role of upper harmonics (see chapter 2). In a highly influential paper, Plomp and Levelt (1965) demonstrated that musically dissonant intervals may be predicted by the concept of the critical band if one considers interactions among upper harmonics. They showed that perceived dissonance is maximal when harmonics of tones fit within the same critical band, while not being identical in frequency. Greatest dissonance is associated with 25 percent of the bandwidth, as shown in Figure 3.1. In contrast, the upper harmonics for highly consonant intervals (e.g., C and G, a perfect fifth) are either identical in frequency or do not fit within the same critical band.

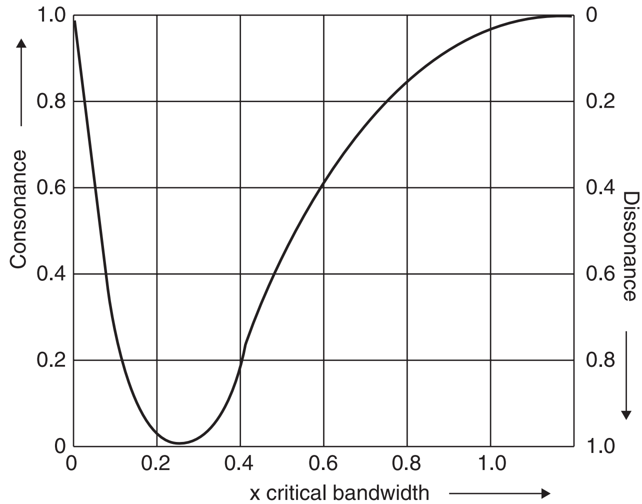


Figure 3.1 Curve representing the subjective consonance/dissonance of two pure tones as a function of frequency separation, shown as proportion of critical bandwidth (Plomp & Levelt, 1965, p. 556).

Source: Copyright © 1965, American Institute of Physics. Reprinted with permission.

Loudness

It may be surprising to learn just how sensitive the auditory system is to differences in the energy of input (that is, in the amplitude of the physical sound wave). In fact, if the ear were more sensitive than it is, we would actually perceive the collisions among air molecules! How loud a sound is perceived to be matters a good deal in music, as in other contexts in which what we hear is of significance. The more energetic the pressure wave – that is, the greater its amplitude – the louder will be the sound heard as the wave impacts on the ear.

The physical loudness of a sound is referred to as its *intensity*. The most common way to represent different levels of intensity is through the decibel scale, which is based on a formal mathematical definition. Specifically: *Intensity in decibels (dB) of a sound is defined as 10 × the logarithm of ratio of the energy of that sound to the energy of faintest discernible sound*. Thus, if we give the faintest discernible sound (defined as having an energy of 10^{-16} watts per centimeter squared) a value of 1, the decibel measure is 10 times log 1/1, that is 10 × log 1. Log 1 is 0, so the faintest sound sets the beginning of the decibel scale as 0 dB. Suppose we double the energy of the impinging sound. The measure will now be 10 × log 2/1, that is log 2. Log 2 is 0.3, so the decibel measure is 3 dB. If we double the energy again, leading to a measure of 10 × log 4/1, we get 6 dB and 10 × log 8/1 gives 9 dB.

To give some reference to ‘real-world’ sounds in decibels: A whisper at a distance of 3 feet away is about 10 to 15 dB, the ambient sound inside a home in a big city is about 40 dB, conversation at a distance of 3 feet is approximately 65 dB, and a hammer hitting a steel plate at about 2 feet away is 114 dB (Pierce, 1992, pp. 119–120). The human auditory apparatus allows people to discriminate loudness differences of approximately 1 dB for frequencies between 200 and 4000 Hz (Yost, 2000). Bats, dogs and other animals have even more sensitive auditory equipment.

Although the decibel scale better approximates perceived loudness than would the raw energy ratio (cf. Moore, 2003), it is very important to note that the decibel scale measures the *physical signal*, not a *perceptual experience*. One's experience of loudness is only roughly proportional to the energy of the impinging sound wave. The physical measure of intensity in decibels does not correlate nicely with perceived loudness in many real-world listening conditions. This is because how loud a sound is perceived to be depends not only on the amplitude of the wave, but also on its frequency. To accommodate this psychological fact into acoustics, S. S. Stevens (1936) introduced the *phon* scale. (For a description of this and other loudness scales, see Rasch & Plomp [1999].)

The US Occupational Health and Safety Administration (OSHA) recommends a threshold of no more than 85–90 decibels over an 8-hour period for safe working environments as prolonged exposure to loud noise can lead to hearing loss. Orchestral performances of Wagner's Ring Cycle opera *Götterdämmerung* reach levels of about 110 dB (quadruple times the OSHA safety threshold!), and rock concerts have been measured at even higher levels. Therefore it is not surprising that the incidence of hearing loss is greater among musicians than nonmusicians (e.g., 42 percent of orchestral musicians exhibit greater hearing loss than expected for their ages, according to Axelsson & Lindgren, 1981). Hearing loss is especially prevalent among those playing the French horn, trumpet, trombone, bassoon, flute, double bass, or percussion instruments (e.g, Axelsson & Lindgren, 1981; Eaton & Gillis, 2002). Damage to hearing is often instrument-specific, with greater loss in the left ear for violinists and in the right ear for flautists (e.g., Ostri, Eller, Dahlin, & Skylv, 1989). Those seated directly in front of loud sections of the orchestra (especially brass or percussion) are also at greater risk for hearing loss. Rock and pop musicians appear to suffer more severe damage, especially in the ear closest to the source of sound such as a loudspeaker or amplifier (Sataloff, 1991).

Listeners are also at risk for hearing loss when listening to music via headphones or attending concerts. For instance, latent cochlear damage is more common in those who frequently use personal stereos than those who do not (LePage & Murray, 1998). Frequent use of earbuds, in particular, may put listeners at risk as the bud fits inside the ear putting the source of sound much closer than any headphone. Most forms of noise-induced hearing loss are permanent, due to irreparable damage to hair cells in the inner ear (discussed later in this chapter). Even short-term exposure to loud music can lead to permanent hearing loss. Park (2003) found maximum noise levels in Karaoke singing environments in Korea to exceed 95 dB to 115 dB, and found up to 8 dB of significant hearing loss (centered at 4000 Hz) in listeners after less than 2 hours of exposure!

Unlike hearing loss due to aging (presbycusis), hearing loss due to exposure to loud music affects the region of frequencies within the musical range. Typically, music-induced hearing loss starts at around 4000 Hz – the high end of the musical range – and gradually moves downward. For instance, due to prolonged exposure to their own singing at levels of up to 100 dB and above, opera singers are at risk for developing a so-called 'auditory scotoma' or 'deaf spot' beginning at around 4000 Hz (Tomatis, 1996). If the scotoma spreads to the 2000 Hz range, the result can be a corresponding loss of richness to the singer's voice due to the inability to hear the higher harmonics. As musicians pride themselves on their ability to hear tonal nuances, this presents a significant loss to both production and reception of music.

Duration

Musical tones also differ in duration. Such differences are represented in musical notation by the symbols for whole note, half note, quarter note, and so on (semibreve, minim, crotchet, quaver, etc.), each representing a tone of one half the duration of the preceding note value. However, these are simply the subdivisions of note durations that conveniently organize tones in most Western music. The ear is sensitive to many other time structures in music – such as the complex meters used in Balkan music, or the rhythmic modes of Arabic and Turkish music – although perceived rhythms often 'migrate' to the closest metric structures familiar to us (see chapter 6 on rhythm). Within Western music, in jazz music and particularly jazz improvisation, and even in the most conservative performances of Bach's music, there are also incremental deviations from strict metronomic time that can still be detected by our sensitive ears. Indeed, the ability to sense these small variations in timing is important to our ability to perceive expressivity in music. (The topic of expressive timing is discussed in greater detail in chapter 11, on performance.)

The duration of the acoustic signal may not solely determine the perceived length of a tone's duration. For instance, when internationally acclaimed percussionist Michael Burritt recorded a series of tones on a marimba – played by either long or short gestures – an acoustic analysis revealed that the resulting tones were acoustically indistinguishable in duration. The long- and short-gesture tones were also indistinguishable to listeners hearing an audio recording of these tones. However, participants tended to perceive the tones to be longer in duration when accompanied by a video showing Burritt playing with long and graceful gestures, and shorter in duration when accompanied by a video showing short, choppy gestures (Schutz & Lipscomb, 2007).

Figure 3.2 shows time-elapsed images of the videos used in this study.

Figure 3.2 Previously unpublished time-elapsed images of the videos used in Schutz and Lipscomb's (2007) study. Captured 200 milliseconds apart, they show that acclaimed percussionist Michael Burritt's striking mallet (held in his right hand) continues moving for a longer time after impact for the 'long' stroke (top row) versus the 'short' stroke (bottom row). Although the difference in post-impact motion does not affect the tone's acoustic duration, it does affect our *perception* of the duration of the tone.

Source: Copyright Michael Schutz and Scott Lipscomb, printed with permission of Michael Burritt.

Notice that the 'long' stroke traces a smooth arc after impact, whereas the 'short' choppy stroke stops abruptly (Schutz, personal communication). Schutz and Lipscomb concluded that the 'difference' participants 'heard' between tones created by long-slow and short-swift strokes was perceptual rather than physical, caused by 'visual artifacts of the performer's acoustically inconsequential gesture' (2007, p. 896). In a subsequent study, Schutz and Kubovy (2009) showed that this illusion persisted even when the video of Burritt was reduced to a point-light display consisting of only a single white dot tracking the moving mallet head against a black background. This is just one of many examples of how the pure physical signal and the perception of a sound may differ. Other examples, pertaining to pitch perception, will be discussed in chapter 5.

Aside from cultural and psychological influences, our perception of tone durations is also circumscribed by the physical limitations of the auditory system. For example, studies of the phenomenon of 'pitch fusion' have shown that chains of acoustical events seem to fuse together into a continuous sound when individual sounds are separated by less than 50 milliseconds (ms). There also seems to be a 'window of simultaneity.' Audible clicks are heard as simultaneous if separated by less than about 2 ms. Similarly there is a threshold of order, in that sound events are heard as separate but their temporal order is not discernible for separations less than 17 ms (Hirsch, 1959). Sounds of very short duration may not even be perceived as a tone. For instance, a very short tone burst (of about 2 to 4 cycles in length) is usually heard as a click without a discernible pitch (Pierce, 1992). The higher in pitch the sound is, the longer the tone burst needs to be if there is to be no click. As we shall see in the following section, there is also a temporal dimension to the perception of timbre.

Timbre

Timbre or 'tone color' is the intrinsic and distinctive quality of the sound produced by a musical source such as a certain instrument or a particular human voice. The brilliant sound of the trumpet contrasts in timbre with the mellow, reedy sound of the clarinet. Timbre perception is a very complex topic that poses many methodological challenges to researchers (see Hajda, Kendall, Carterette, & Harshberger, 1997). For the purposes of this brief discussion, we focus on two aspects: *frequency content* and *changes across time*.

The quality we refer to as 'timbre' corresponds with the complexity of the waveform, which reflects the various frequencies it contains. A pure tone, such as the ringing of a tuning fork, is produced by a sound wave of a single frequency of vibration. However, as discussed in the previous chapter, the sounds produced by most musical instruments are comprised not only of the fundamental pitch but its upper harmonics or overtones. The particular combination of overtones sounded by an instrument is responsible, in part, for its distinctive tone quality or 'timbre.'

Figure 3.3 shows the characteristic waveforms of a guitar, violin, vibraphone, and double bass, along with associated frequency spectra. The waveforms (left column) are best used to see differences in fundamental frequency (pitch); in complex tones like the ones shown here, fundamental frequency relates to the repetition rate of the overall complex vibration pattern. The double bass plays a low pitch, and you can see just three repetitions of its vibration pattern, whereas the vibraphone and violin play higher pitches with associated oscillations that repeat many times in the span of 50 ms. The guitar falls in the middle with 10 repetitions. Frequency spectra (shown on the right) relate mostly to differences in timbre. The frequency spectra of the two instruments displayed on the top (guitar and violin) have more energy at high frequencies than do the two instruments displayed below. These differences reflect the timbre of instruments – whereas the violin can have a more piercing sound, the vibraphone is more 'mellow' and the double bass is richly resonant. The guitar greatly resembles the violin – in fact the difference in timbre between guitar and violin (if both are plucked) is due mostly to a single post inside the body of the violin that supports the bridge holding the strings (Handel, 1989).

Along the temporal (i.e., time) dimension, the 'envelope' of a non-percussive musical signal can be divided into three segments: *attack*, *steady state*, and *decay*. The 'attack' (roughly, onset to initial achievement of a steady tone) of a bowed violin is quite slow; it takes 1/5 of a second or more before the string is fully engaged in the regular motion that produces the 'steady state' (Beament, 1997). By comparison, a cymbal struck by a hard mallet has an almost instant attack, no true steady state, and a very slow decay. The attack may play an important role for the identification of musical instruments (though see Hajda et al., 1997, for further discussion). For instance, if the initial portion of a recorded tone (such as the first 1/5th of a second of the starting sound of a trumpet) is removed, it is often more difficult for listeners to identify the instrument (Saldanha and Corso, 1964). This may be in part because instruments have distinctive *attack transients*, or temporary fluctuations in the sound signal that occur during the initial phase of a tone. Attack transients are largely responsible for the starting 'blat' of a trumpet and the initial gritty sound of a low cello tone. This is not to say that steady state and decay are not also important for identification of musical instruments. For instance, Saldanha and Corso found that instruments with a vibrato tone in the steady state were more easily identified than were those with a non-vibrato tone. Further, while transients seem to provide important cues for identifying the timbre of single musical tones, Kendall (1986) found the steady state to be more important when identifying instruments playing short musical phrases.

Recently, researchers have begun to investigate whether computers can be programmed to identify musical timbres. While there is a large body of work on speaker identification by computers, few studies have focused on

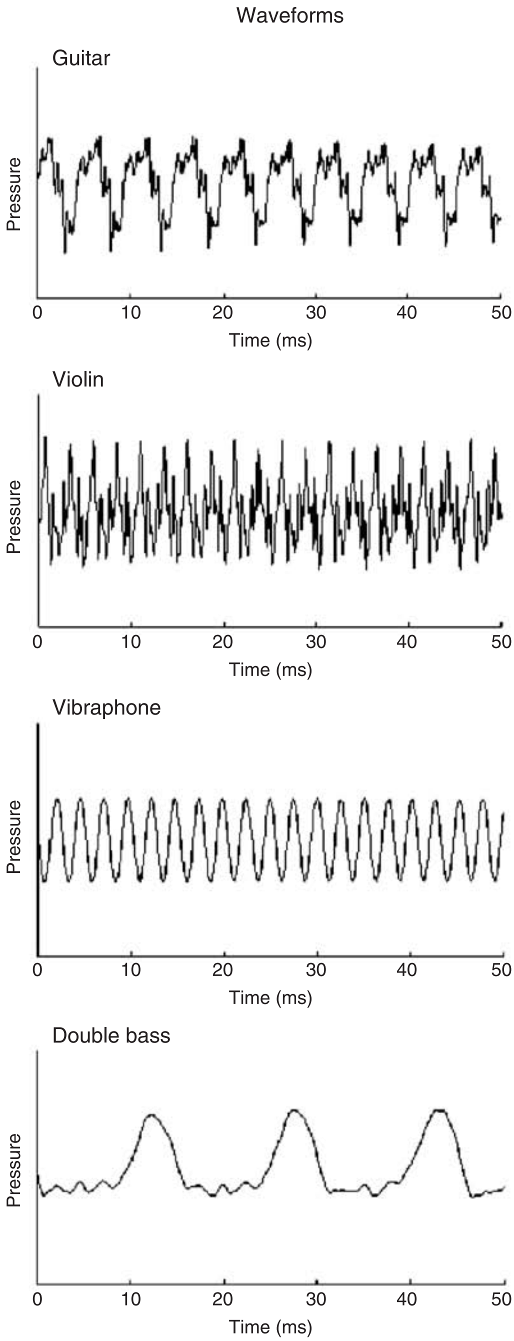
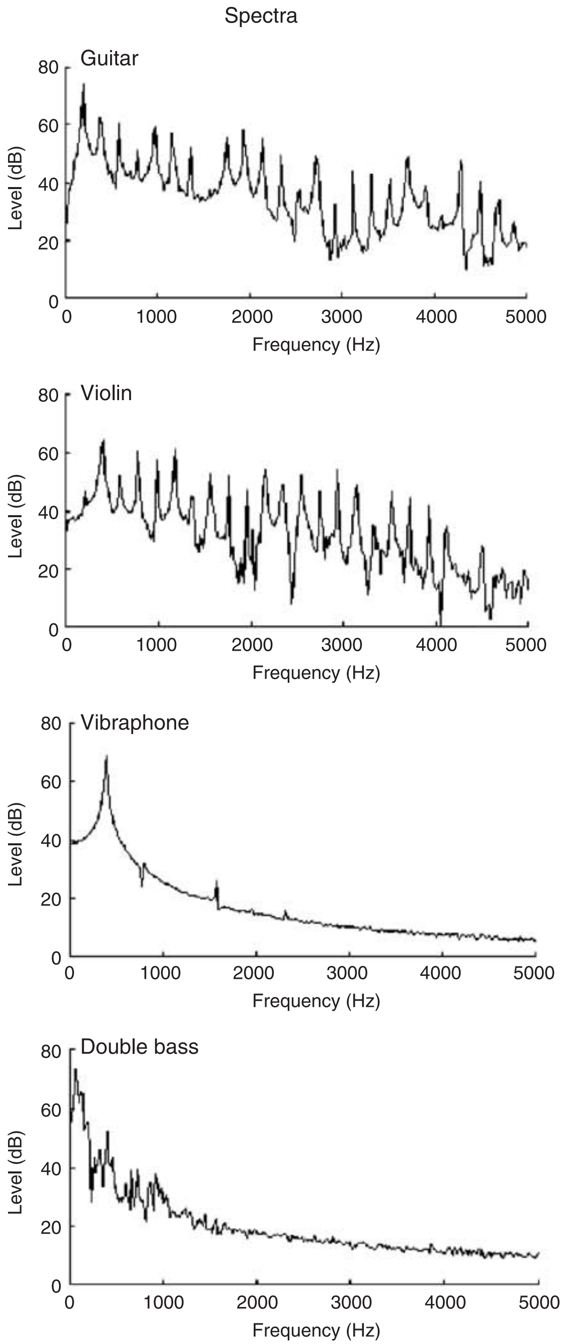


Figure 3.3 Waveforms (left) and power spectra (right) for a variety of musical instruments.

Source: Reproduced from Plack (2005, p. 41) by permission of the author.


Spectra
Guitar
Violin
Vibraphone
Double bass
Level (dB)
Frequency (Hz)
0
20
40
60
80
0
1000
2000
3000
4000
5000

identification of musical instruments. In one study, Brown (2001) found that computers performed comparably to humans in identifying four wind instruments. Flute, clarinet, oboe, and saxophone were chosen as these timbres are often challenging for humans to identify. It was notable that Brown used excerpts from musical recordings by professional instrumentalists so that the tones were taken from natural performances in musical context rather than single pitches produced in the laboratory.

In chapters 2 and 3 we have discussed various properties of sound and how sound is perceived in a musical context. The function of the auditory system is to serve as the neural basis of the processes that bring about this ultimate transduction, from the physical properties of material vibrations to patterns of electrical and chemical events in the neural nets of the brain to the sounds we hear as music. However, the complexity of the neuropsychology of hearing, particularly when the topic is the perception and production of music, makes it desirable to describe the anatomy and physiology of the auditory apparatus, the outer, middle and inner ear, before turning to the processes in the brain through which auditory experience is produced.

The anatomy and physiology of the ear

The auditory system consists of the external ear, the middle ear, and the inner ear. Its major divisions are shown in Figure 3.4. The *external ear* gathers the trains of pressure waves in the air and directs them into the auditory canal. The *middle ear* includes the apparatus of tiny bones, or ossicles, by which sounds are transmitted to the *inner ear*, the main organ of which is the cochlea. The cochlea is the wonderful device that, with its complex sound detectors, serves to transduce the relevant physical properties of wave trains into neural impulses, which are transmitted through complex networks of neural tracts to various parts of the brain.

The external ear

The external ear is comprised of the outer ear (the *pinna*, or the ear 'flap' that protrudes from the head) and the *auditory canal*. The pinna is the device that 'scoops up' the sound waves, directing the incoming vibrations into the auditory canal toward the eardrum. The auditory canal, an S-shaped tube of about 2.5 cm in length, terminates in the *eardrum* or *tympanic membrane* (the membrane that vibrates to the energy and frequencies of the incoming wave trains). The auditory canal acts as a resonator (especially for frequencies ranging in the highest octave of a grand piano), so there is some amplification of the sound waves as they travel down the ear canal.

The *pinna*, or *outer ear*, plays an important role in detecting where sounds are coming from, especially for discriminating whether sounds originate in front of or behind us. The asymmetrical form of the pinna (i.e., that it is a half-moon shape, joined on one side to the head) serves us well in this respect.

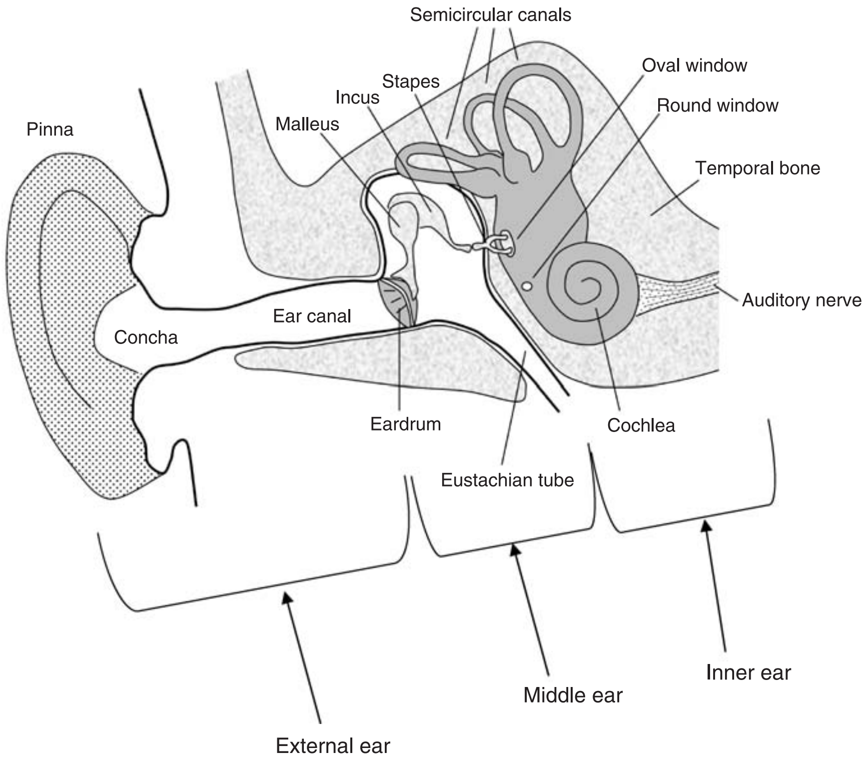


Figure 3.4 The main structures of the ear.

Source: Reproduced from Plack (2005, p. 64) by permission of the author.

If the ear were completely round and affixed to the head by its center, localization of sounds originating in front or behind us would be extremely difficult! Further, it is the irregularity of the convolutions of the outer ear that offers important cues about the location of a sound, as the convolutions create a series of reflected waves before the sounds enter the ear.

The shape of the pinna varies widely in the population, so sound localization cues that rely on the pinnae are highly idiosyncratic and must be learned. Indeed, when four participants were fitted with small prostheses that changed the shape of their outer ears and wore them continuously for six weeks, their ability to localize sounds accurately declined! By about six weeks, their performance rose again to baseline levels as they grew accustomed to their 'new ears' (Hofman, Van Riswick, & Van Opstal, 1998).

Unlike most mammals, human beings cannot move their pinnae toward the direction of sounds; they must move the whole head to achieve this task. Owls, in particular, are extremely accurate at localizing sound sources in part because of their capacity to rotate the head by about 270 degrees (not exactly a complete rotation, as commonly thought). Further, the entire facial ruff (the long feathers on the head and neck) serves a similar function to the human pinna, as these sound-reflecting feathers provide cues as to the source of an auditory event (Knudsen & Knudsen, 1989).

Of course, sound localization does not rely on the outer ear alone. We also depend on *binaural* cues such as the interaural differences in time of arrival and intensity of a sound to each ear (in addition to *monoaural* cues supplied in part by the pinnae, as described above). In other words, because we possess two ears that are spaced some distance apart, a sound originating from a source closer to the right ear arrives at the right ear slightly faster (*interaural timing*) and/or at a higher intensity (*interaural intensity or level*) than at the left ear. These and other cues (such as the angle at which the sound waves hit the outer ear) are helpful in localizing a sound. While these cues do explain source detection under many conditions, sound localization appears to be somewhat more complex. The entire body, especially the head, shoulders, and chest, plays a role in reflecting and diffracting sound waves, bringing about significant changes in the composition of frequencies which reach the ear. Another factor may have to do with the fact that the phase, that is the pattern of the crests (high-pressure phase of the sound wave) and troughs (low-pressure phase of the sound wave) of the incoming waves, are different at each ear unless the head is aligned with the track of the sound waves. A full discussion of this effect is covered in volumes focusing on auditory perception (e.g., Handel, 1989, ch. 3).

The eardrum and the transmission system of the middle ear

The middle ear consists of a system of three tiny bones (or *ossicles*) that directly link the tympanic membrane with the *oval window* of the cochlea. These ossicles are illustrated in Figure 3.4 and are referred to as the *malleus*, *incus*, and *stapes* (or *hammer*, *anvil*, and *stirrup* respectively, named for their distinctive shapes). The malleus is attached to the inner surface of the tympanic membrane, and is connected to the incus by a flexible joint. The incus is connected to the stirrup-shaped ossicle, the stapes. The stapes ends in an oval 'footplate' which is fixed to the oval window of the cochlea.

The middle ear is designed to preserve information in the sound wave as it is transferred from one medium (air) to another (fluid). The problem that the middle ear is designed to solve is that fluid is of higher density than air, and so a similar amount of energy will be less effective in moving fluid than in moving air (thinking about waving your arms in the air versus through water). Thus the middle ear accomplishes *impedance matching*: that is, the preservation of energy across context characterized by differing levels of resistance. The *impedance* of any medium is defined as the ratio of the incident pressure to the flow thereby induced in the medium. The air – the medium of transmission for sound in the external environment – is of low impedance; that is, it has little intrinsic resistance to motion. However, the cochlea is filled with fluid, which is of high impedance. Remember that a sound wave involves the movement of the molecules in the medium along the line of transmission (chapter 2). So, how is the motion of air molecules faithfully transmitted to molecules of the liquid in the cochlea, given the greater impedance of the cochlear fluid? The ossicles perform this function. Though the total energy of the sound wave is maintained across the ossicular mechanism, the amplitude of the wave is increased.

The middle ear is also designed to keep loud sounds out of the cochlea. Two muscles (*tympanic* and *stapedius*) attached to the ossicles contract in response to very loud sounds, attenuating low-frequency sounds below about 1000 Hz. This *acoustic reflex* protects us from loud incoming sounds but also occurs when we talk or sing (even at moderate levels of 65 or 75 dB), shielding us from the effects of prolonged exposure to our own voices. The reflex, however, is too slow to protect the inner ear from sudden or intermittent loud sounds (which is why ear-plugs are essential for percussionists).

Initial research on the amplification function of the middle ear was conducted by Georg von Békésy (1899–1972), a Hungarian-born physiologist working in the second quarter of the twentieth century. While employed in the research laboratory to improve telecommunications at the Hungarian Post Office, von Békésy became interested in the functioning of the ear as the essential component in the transmission of signals. He embarked on an extensive program of study, conducting postmortem examinations of the internal structures of the ears in humans and a wide range of animals, from the tiniest mouse to the great elephant. Reportedly, he often used the drill press at the Post Office where he worked to drill holes in the ears of cadavers! He also pioneered an ingenious method of sprinkling silver flakes on the basilar membrane and using strobe photography to examine the way it moved in response to different tones. In many areas of research, and specifically in the investigation of how the basilar membrane vibrates in response to sound, von Békésy continued the work of Hermann von Helmholtz (chapter 2).

Von Békésy's postmortem studies of the human middle ear have remained the major source of our knowledge about the functioning of the complex mechanism of the ossicles. His many papers have been collected in *Experiments in Hearing* (von Békésy, 1960). However, the work for which he received the highest recognition was his research on the functioning of the cochlea. In 1961, he was awarded the Nobel Prize for Physiology or Medicine, for his 'discoveries of the physical mechanisms of stimulation within the cochlea.' It is to this amazing organ of the inner ear that we turn our attention next.

The inner ear: The structure of the cochlea

The main structure of the inner ear is the cochlea. Figure 3.5 shows the cochlea as if unrolled, viewed from the side. The *cochlea* is a snail-like arrangement of three parallel tubes – the *vestibular canal* (*scala* is used for canal in the figure), the *median canal* and the *tympanic canal* – filled with an incompressible fluid (*endolymph* in the median canal, *perilymph* in the

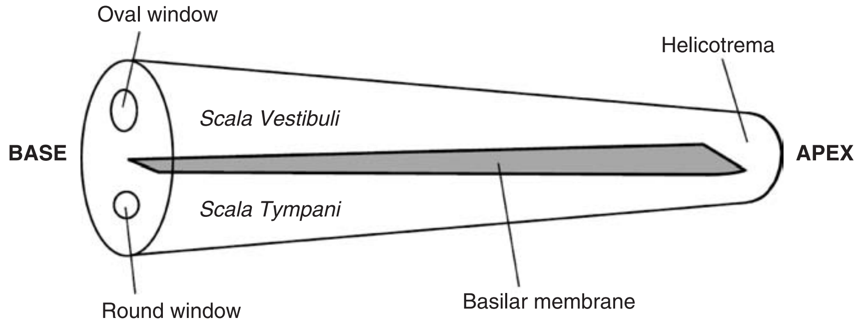


Figure 3.5 The structure of the unrolled cochlea.
Source: Reproduced from Plack (2005, p. 66) by permission of the author.

others). If the cochlea were unrolled (as in Figure 3.5) it would measure about 3.5 centimeters in humans. The median canal is separated from the others by two flexible membranes, the basilar membrane (which separates median from tympanic) and Reissner's membrane (which separates median from vestibular). The median canal contains structures that allow auditory perception to occur, and we focus on these structures next.

On the surface of the basilar membrane, within the median canal, the entire neural apparatus for sound detection is laid out in the *organ of Corti*, shown in Figure 3.6. The perspective of Figure 3.6 is a 90-degree rotation of the perspective used in Figure 3.5, as if we were peering down the length of the cochlea, and the level of detail is much higher, as we are focusing on a specific structure in the median canal – a mere sliver in Figure 3.5. The organ of Corti contains bands of sensitive *hair cells*, in two rows, laid out in an inner (shown on the left in Figure 3.6) and an outer array (shown on the right).

How does sound travel through the ear? Pressure is transmitted by the stapes, fixed to the flexible membrane in the *oval window* of the vestibular canal. This results in pressure waves that are induced in the fluid in the cochlea by the action of the stapes on the membrane of the oval window. Since the vestibular fluid is incompressible, the pressure of the wave must be relieved. This is achieved by the flexing of the membrane in the *round window* (the second aperture in the tympanic canal), when the pressure wave reaches it. (The main structures described above are labeled in Figure 3.5 and in our main illustration of the ear in Figure 3.4.) The main mechanism of pitch detection depends on the induction of a traveling wave in the basilar membrane. In order to understand this process we need to know more about how the cochlea both *transduces* and *encodes* frequency information in the sound wave.

Neurophysiology of the cochlea as transducer

A complex process occurs during sensation: energy of one form (like a sound wave) must be converted into an electro-chemical energy (a neural impulse).

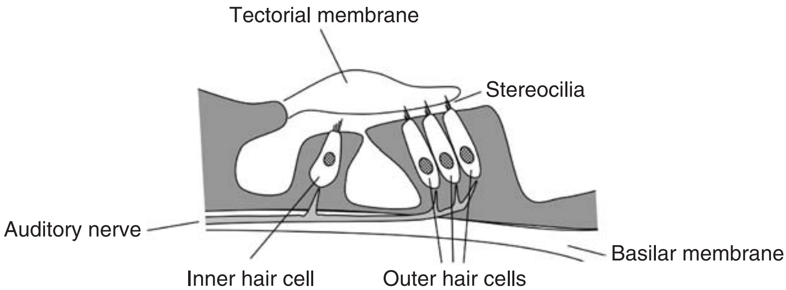


Figure 3.6 A cross-section of the cochlea rotated 90 degrees, showing the organ of Corti, including inner and outer rows of hair cells.

Source: Reproduced from Plack (2005, p. 67) by permission of the author.

This process is called *transduction*. This transformation must preserve certain *relevant* features of the original process. For instance, when pressing keys while typing at a computer keyboard, we are involved in the process of transducing mechanical impulses (movements of keys) to electrical signals in the device. The device has to retain information about what key was pressed even though the physical signal has changed significantly.

When we hear music, the pattern of physical disturbances from the original vibration (e.g., vibrating vocal cords, in the case of singing) is generally retained through the ear. But following the wave's arrival in the cochlea, transduction must occur in order for the information to reach the brain, since the brain only 'understands' electro-chemical signals. For music, the relevant physical information can be reduced to frequency (heard as pitch) and amplitude (heard as loudness). The form of this information, prior to neural coding, is that of induced pressure waves in the perilymph surrounding the median canal. The pitch and amplitude of the initiating sound wave must find some 'representation' in the pattern of neural impulses exiting the right and left cochleas. Somehow, all the energy resulting from colliding fluid molecules in the inner ear must be converted to neural signals that are reconstructed in the brain so that we can experience the intended musical message being performed.

As the basilar membrane is progressively displaced by the movement of fluid in the cochlea, hair cells embedded in the *organ of Corti* (Figure 3.6) push against the dense and positively charged endolymph that fills the median canal. When this happens, hair cells bend and a small opening at the base of the hair cell is opened, and positively charged potassium ions flow inward. This alters the potential difference between the inner and outer regions of the hair cell, thus creating an action potential. At this point, sound energy is transduced into a nerve impulse, though only the inner hair cells convey information about sound vibrations (outer hair cells serve a different purpose not addressed here; see Pickles [2008]). A complex tone with three frequencies of 200 Hz, 400 Hz, and 800 Hz, for example, will lead to a traveling wave in the basilar membrane with three regions of relatively maximal oscillations. As a result, sets of hair cells at three corresponding places on the membrane will respond, thus mapping sound frequency onto the cochlea tonotopically (as will be explained in the next section).

Encoding of pitch in the cochlea

Transduction involves a dramatic change in the basic format that information assumes. It is important that this change in format does not lead to a change in content. That is, a sensory system must preserve information from the outside world even as that information is transduced. Such preservation of information is called *encoding*. The cochlea is designed elegantly to encode information about frequency in sound; an important issue for the perception of music which we discuss briefly here.

Our understanding of the neurophysiology of the cochlea is based on the results of the extensive research program undertaken by von Békésy. From von Békésy, we learned that areas of the cochlea respond differentially to sounds depending on their frequencies. Since then, two dominant theories have emerged to explain how the frequency of a sound wave is encoded as pitch, one for sounds of frequency more than 5000 Hz and one for sounds of lower frequencies. These are referred to as the place theory and time theory for the encoding of pitch. Both theories ultimately seem to be necessary to account for the perception of pitch (Cariani & Delgutte, 1996).

According to the *'place' theory*, the relevant parameter is *where* on the basilar membrane the maximum displacement occurs. This location will be populated by a certain row of inner hair cells which will be maximally stimulated by the displacement of the membrane, perhaps amplified by changes in the shape of the corresponding outer rows of hair cells. The place theory suggests that the basilar membrane responds in a way that segments the components of a complex tone, performing a sort of Fourier analysis. In such a framework, pitch perception may be determined by the 'place' associated with the fundamental frequency. However, place mechanisms have long experienced difficulty explaining the aforementioned illusion of '*residue pitch*,' for which there is no 'place' corresponding to the fundamental that is heard as the pitch. This problem has led some to propose that pitch perception results from temporal properties in the overall spectrum of a complex tone, as in the theory we describe next.

According to the *'time' theory*, the relevant parameter is the *time* between successive phases of the displacement of the basilar membrane by the energy of the traveling wave. A sine wave makes a complete cycle from 'crest' to 'trough' and back again to a 'crest.' The higher the frequency, the shorter the time between each half phase of the cycle, and so the more rapidly the bursts of firings in the auditory nerve will follow one another. With respect to complex tones, the overall pattern of firing can be used to identify the fundamental even if the fundamental frequency is absent and one must rely on residue pitch instead (cf. Cariani & Delgutte, 1996). However, the time theory also runs into problems. For vibrations of less than 5 kHz the auditory nerve bundle fires for only half the cycle of stimulation, but for vibrations above 5 kHz it fires continuously. If the auditory nerve fires continuously, this source of pitch information disappears since there is no time difference between successive levels of neural activity, for example successive 'crests' of the energy wave. Thus the time and place mechanisms may be suited to different kinds of limitations on our ability to perceive pitch. Timing information is the major factor in encoding pitches below 5 kHz and place information for pitches above 5 kHz (Pickles, 2008, p. 273).

So far, we have described in outline how the outer, middle, and inner ear play their parts in an auditory transducer. Pressure waves in the air are represented by impulses from the neural apparatus of the pair of cochleas. These impulses pass into the brain, and are eventually experienced by listeners as sounds, and some of these sounds as music. The next step will be to describe the findings of research into the relevant workings of the brain.

Auditory pathways to the brain

The neural pathways from cochlea to auditory cortex

The pathways leading from ear to cortex are complex and tangled. Though a great amount of physiological research has been done on these connections, the significance of these findings for the experience of complex sound patterns such as music is not yet clear. A remarkable theme in these pathways, in contrast to visual pathways, is the degree to which auditory pathways provide both *ipsilateral* (same side) and *contralateral* (opposite side) connections from ear to brain. In other words, there is much more sharing between the left and right in our auditory system than in our visual system. Some neurons receive excitatory input from one side and inhibitory input from the other side (Plack, 2005). This kind of setup may relate to the fact that spatial relationships in hearing often require a comparison between inputs to the left and right ears, such as when listening to music through stereophonic headphones.

The auditory cortex

The *auditory cortex* is a bilateral structure in both left and right temporal lobes, consisting of three layers of neurons, the *primary*, *secondary*, and *tertiary* regions. The layers are arranged in concentric circles, the primary auditory cortex occupying the center surrounded by the cells of the secondary and tertiary regions. The *primary auditory cortex* consists of columns of neurons that appear to be *tonotopically* arranged; that is to say each cell in a column is tuned to the neuronal signal from a particular frequency in the auditory input into the cochlea. Tonotopical mapping matches a location in the primary auditory cortex to a location on the basilar membrane of the inner ear. Thus the organization of the primary auditory cortex matches that of the tonotopical layout of the bands of hair cells along the basilar membrane in the cochlea. Higher frequencies are recognized towards the central area of the primary auditory cortex and lower frequencies nearer the boundaries. The significance of tonotopic organization is that the auditory system represents pitch in a manner analogous to a piano, in that there is a spatial representation of pitch. Furthermore, the tonotopic organization is laid out such that the brain encodes pitch in absolute terms (e.g., there is a specific region associated with C4 or the fourth highest C on a grand piano). However, if primary auditory cortex is organized in absolute terms, the ability to perceive pitch in this way, called 'absolute' or 'perfect' pitch ability, should be more widespread than it is. We discuss this interesting topic further in chapter 5.

Adjacent to each primary auditory cortex are the *secondary* and *tertiary auditory regions* in concentric layers, referred to as the '*belt*' and '*parabelt*.' These are illustrated in Figure 3.7. In this figure, the fold containing the auditory cortex (called Heschl's gyrus) has been cut and opened. This figure shows the brain of a macaque monkey, which is structurally similar to that of a human brain with respect to the auditory cortices. As their names imply, these regions appear to be important for more complex aspects of auditory experience. Experiments have shown that the cells in these regions do not respond at all, or if they do so, only weakly, to neuronal signals when pure tones are being detected by the cochlea. By contrast, these areas respond to complex tones in ways that distinguish 'what' from 'where' similarly to higher processing areas of the visual system (Rauschecker & Tian, 2000). This may have practical implications for music listening in that, for instance, we can

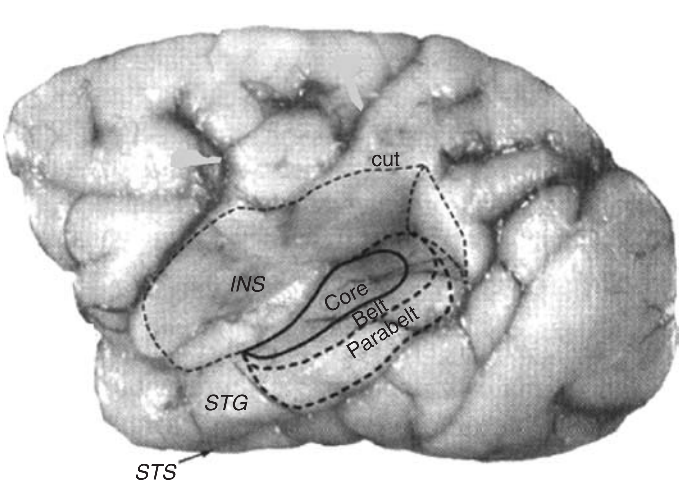


Figure 3.7 Secondary and tertiary auditory regions in the macaque monkey (Kaas & Hackett, 2005, p. 11). Abbreviations in the figure are *INS* = insula, *STG* = superior temporal gyrus, *STS* = superior temporal sulcus.

Source: Reprinted with permission from Lawrence Erlbaum.

identify the sound of an instrument independently of its location in space (e.g., as a solo singer walks across the stage).

Coda

Somehow the qualities of sound heard as music must be extracted from compression waves propagated through the air from a source to a person who not only hears sounds, but music. To mediate the transduction from movements of air molecules to musical experience, there must a transducer – a device that reproduces the energy pattern in one medium into an energy pattern in another. This device is the cochlea. The system involves four parts: a receiver for pressure waves in the ambient air (or water), the pinnae; an amplifier that transmits differential pressures to the basic transducer, the eardrum and auditory ossicles; the transducer itself, the pair of cochleas; and finally the pattern of nerve fibers that collects signals from the cochleas and transmits them to the cross-connected paired auditory cortices.

The brain is the organ we use for every imaginable cognitive task, including the perception of sound as music. Before this can happen, the incoming sound wave must be analyzed for its musically relevant properties, particularly amplitude and frequency. Again, this is made possible by the role of the cochlea as transducer. The activation of the pressure-sensitive hair cells in the spiral tubes of the cochlea initiates a very complex process by which both the fundamental frequency of a note and its harmonics are abstracted as neural inputs into the auditory cortex. Thanks to the extensive studies of von Békésy, we now know how this remarkable organ is able to perform this analytical task. It can also be said with some confidence that the workings of the primary auditory cortex in distinguishing sounds by pitch relations have also been worked out. Having summarized the basic neurophysiology of hearing, we now turn to the broader neural bases of music perception and performance.

4 Neuroscience and music

New technologies often bring great hope. Such has been the case during the last couple of decades or so with the advent of brain-imaging technologies. These new techniques allow us to see the inner workings of the brain, and may even eventually let us watch the process of thought itself. However, the brain is an immensely complex organ; the number of neural connections in a single person's brain has been estimated to equal the number of atoms in the universe! Needless to say, new technologies are just beginning to help us understand how the brain works, but in the least these advances offer a window into the intricate and mysterious workings of the mind.

This chapter serves as a general introduction to cognitive neuroscience, both conceptual and methodological, as is relevant for its application to the psychology of music. *Neuroscience* is a broad term that includes many subfields, all of which focus on the brain. It includes traditional *neurology* (the study of patients who suffer from brain damage) and *neurophysiology* (the study of the nervous stem on a fine-grained level, often cellular), whereas *brain imaging* is more commonly associated with modern techniques such as PET (positron emission tomography) and fMRI (functional magnetic resonance imaging). *Cognitive neuroscience* refers to the application of all these neuroscience techniques to understand the biological bases of thought. This chapter will focus primarily on brain imaging, but will also draw from the other subfields. We begin with a consideration of a few conceptual issues pertinent to neuroscientific research, followed by a brief introduction to the main structures of the brain and description of neuroscientific methods. This knowledge is extended by a discussion of recent research addressing possible links between music and language.

The first sections of the chapter are intended to provide a basic working knowledge of the main structures of the brain and some methods used in this field. Those with a background in these areas may wish to proceed straight to the discussion on the possible links between music and language. Our discussion of the neuroscience of music will be continued in five topics in this book: the perception of melody and rhythm (chapters 5 and 6), music practice and performance (chapters 10 and 11), and emotion in music (chapter 14).

Three 'big questions' for the neuroscience of music

It is easy to get lost in the details of neuroscience and forget the larger questions that motivate this approach. As such we begin this chapter with a broad perspective, by touching on three fundamental questions that have important implications for research in this field. These are 'large' issues that require extensive discussions that lie beyond the scope of this chapter, so our aim here is to provoke thought rather than provide answers.

Genes or environment? Perhaps the most fundamental question one comes across in a discussion of human behavior concerns the role of 'nature' or 'nurture' in a phenotype. In this context, the term 'nature' refers to the influence of genetics, whereas 'nurture' refers to the impact of the environment. This question is difficult to resolve because both factors contribute to the thoughts, behaviors, and abilities we exhibit and – importantly – both guide the development of the brain. Nevertheless, the question of nature versus nurture is of real relevance to researchers interested in musical abilities, as evidenced by discussions about whether musical 'talent' exists (e.g., Ericsson, Krampe, & Tesch-Römer, 1993; Howe, Davidson, & Sloboda, 1998). It is tempting to think that 'biological' evidence from brain-based research represents the 'nature' side. However, drawing such a conclusion is hasty because, as we shall see, the brain changes with experience. Further, we know now (since the mid 1990s) that certain regions in the adult brain even regenerate brain cells. Certainly biological evidence can address the nature/nurture debate, but one must look to genetic or comparative (i.e., cross/species) evidence, not merely to the workings of the human brain.

To what extent are certain behaviors associated with particular areas of the brain? It is common in this literature to read suggestions that a certain cognitive characteristic (e.g., pitch perception) is governed by neural tissue at a certain location (e.g., primary auditory cortex). The assumption underlying this view is referred to as *modularity* (Fodor, 1983): the idea that large-scale, complex qualities of the mind (e.g., memory) result from the summed activity of various component processes, each of which functions independently of the other and accomplishes relatively simplistic tasks with great efficiency. An alternative view, held by many, is that the brain functions like an integrated network of connections. This view is often thought of as the *connectionist* perspective (McLeod, Plunkett, & Rolls, 1998).

Are there regions that are 'dedicated' to music alone? In practice, it can be difficult to determine with certainty whether there are fixed links between brain structures and behavioral functions. Often it is thought that structure X governs function Y only because it is not yet known that structure X also governs function Z. As we will see, neuroscientists' views on which brain areas are 'specialized' for language have changed over time. Moreover, people have come to see the relationship between brain and behavior as more flexible than had been previously thought (e.g., McMullen & Saffran, 2004).

Does studying the brain really reveal to us how the mind works? It is almost taken for granted in mainstream neuroscience that the brain and mind are the same. By extension, one might assume that thinking is nothing more than brain activity. With respect to music performance, for instance, this idea would imply that a performance that moves you to tears is ultimately the result of a chain of neural reactions, and can ultimately be understood in those terms. The emotional response of the listener (tears) likewise is understood as the brain's response to sensory input from the performer. It is worth pointing out that consensus on this issue has not always existed, and even today many thinkers are skeptical of the view that mind and brain are simply one and the same.

It is important to recognize the philosophical precept that guides current mainstream neuroscience, referred to as the *mind–brain identity thesis*, which is a brand of *monism*. This view of the nature of human beings is of great antiquity. Whereas *monism* stipulates that there is only one substance – whatever that is – the mind–brain identity thesis goes on to suggest that this single substance may have two different kinds of attributes: 'material' and 'mental.' In recent times this view has been called 'neutral monism' because there is no fixed commitment to the substance of thought being either material (brain) or mental. Among philosophers this conception of the nature of human beings was most fully worked out by Baruch Spinoza (1632–1677), and among psychologists by Antonio Damasio (1994). Spinoza (1677) called the two aspects of the basic substance 'modes.' Following this view, the basis for mental life resides in a single entity. Considered one way, say by a neurosurgeon, it performs material functions, the topic of neuroscience. Considered another way, its activities are experienced as thoughts and feelings. Moreover, though the electrical and chemical activity in the brain and central nervous system has attributes that are radically distinct from the attributes of thinking processes, they are each attributes of just one substance. Feeling miserable is one way that the common activity of the human body can be viewed; excessive serotonin reuptake at the synapses is another.

Neural structure and function

The gross anatomy of the brain

We now introduce the basic structures of the brain, focusing in particular on those structures that figure prominently in the research on the experience and performance of music. Throughout Europe until very recently, every educated person had some knowledge of Latin and Greek. Brain structures, like most other 'objects' of interest to scientists, were given names derived from those languages. In brain nomenclature from the earliest days, the terminology was based on Greek, since the most important work was performed by Greek physicians such as Galen, or translated into Greek from Arabic. For instance, an important region of the brain is the *hippocampus*, from 'hippo' meaning horse and 'kampi' meaning curved (or 'sea horse' in Ancient Greek). Thus this structural feature, which looks something like a sea horse, became the 'hippocampus.' Further, different subfields (e.g., neurophysiology versus brain imaging) often use different terminological systems. Thus we offer a brief introduction to some basic terminology that will be helpful in reviewing much of the literature.

In our discussion of music neuroscience we focus primarily on structures of the *neocortex*, the most recently evolved section of the brain, which forms the surface of the brain as one sees it excised from the organism's skull. The most common way to identify areas of the neocortex is to categorize them in *lobes* and *hemispheres*. The hemisphere distinction relates to the right versus left side of the brain. Hemispheres are structurally separate but are for the most part *homologous* to each other, at least when it comes to gross anatomy (such as the arrangement of lobes). That is, structures in the left hemisphere are replicated in the right hemisphere. Figure 4.1 shows the right hemisphere, and illustrates its lobes, which are also present in the left hemisphere. Lobes are not designated according to neural function or neuroanatomy; they are distinguished simply from the bones that surround the brain (that is, the *frontal, parietal*, *temporal*, and *occipital* lobes are named for the corresponding cranial bones that underlie them). The notion that each lobe performs a specific task is only partly true. Ultimately most of the brain is involved in

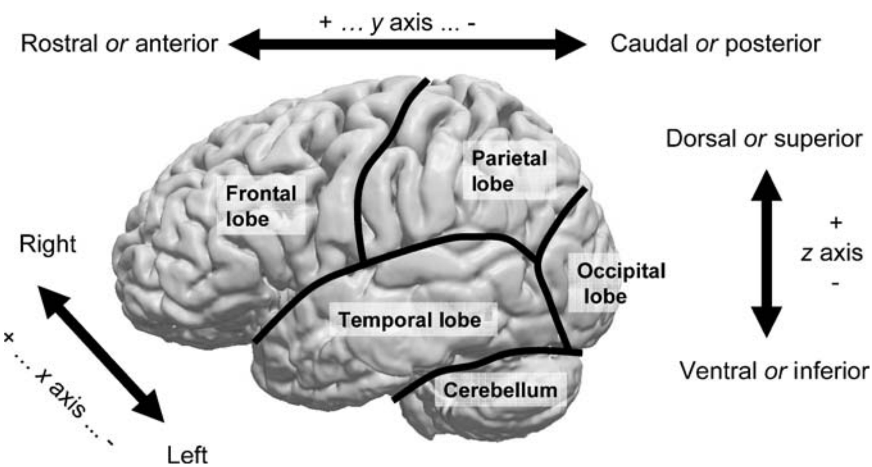


Figure 4.1 Lobes of the neocortex and the cerebellum, along with coordinate labels. The left hemisphere of author PQP's brain is shown.

Source: Brain image scans produced in collaboration with the Buffalo Neuroimaging Analysis Center (Robert Zivadinov, M.D., Ph.D., Jennifer L. Cox, Ph.D.) www.bnac.net. Used by permission from BNAC.

musical behaviors, which is typical of complex activities. In this book we will focus largely on the temporal and frontal lobes, based on their dominant role in hearing and motor actions, respectively.

It is often useful to refer to brain structures and their locations relative to one another (where is region A relative to region B?), rather than simply by spatial location. The terms used for relative locations in the brain were originally designed with a four-legged creature in mind. This becomes problematic because a quadruped standing on all fours positions its head so that the jaw is roughly parallel to the trunk, whereas it is closer to perpendicular for us. Thus, the term for brain regions toward the bottom are called 'ventral' referring to the 'belly-side' and brain regions toward the top are called 'dorsal' referring to the 'back.' 'Dorsal' areas are sometimes referred to as 'superior' to (meaning 'above', not 'better than') those that are below them. Areas toward the front of the brain may be called 'anterior' or 'rostral' (or sometimes simply 'frontal'), and areas toward the back of the brain may be called 'posterior' or 'caudal.' Again, consistent with the stance of a four-legged creature, 'rostral' refers to the snout or beak and 'caudal' to the tail. Areas toward the left or right are simply called 'left' or 'right,' and areas toward the center are referred to as 'medial.'

Figure 4.2 indicates some brain regions that figure prominently in research on music and the brain. Note that every brain is different and different studies lead to different conclusions about exactly where musical 'activations' occur.

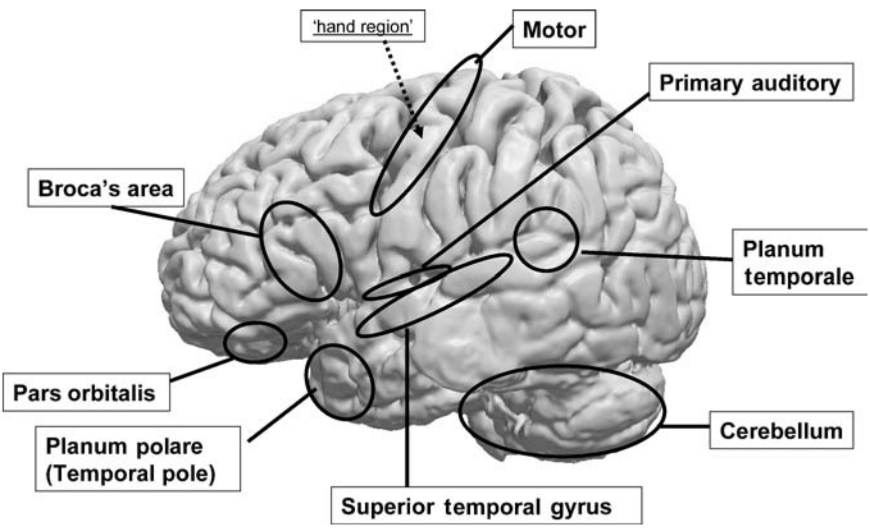


Figure 4.2 Surface regions of the brain associated with music. The left hemisphere of author PQP's brain is shown. Arrows are used to indicate regions on a more localized scale.

Source: Brain image scans produced in collaboration with the Buffalo Neuroimaging Analysis Center (Robert Zivadinov, M.D., Ph.D., Jennifer L. Cox, Ph.D.) www.bnac.net. Used by permission from BNAC.

Regions highlighted in Figure 4.2 should thus not be considered as absolutely precise but rather as gross approximations to the actual brain areas that figure in musical behaviors. You can see that the primary areas used in music are areas associated with hearing (superior temporal lobe) and motor control (caudal and ventral frontal lobe, as well as the *cerebellum*). The cerebellum, which is an area of the hindbrain, is sometimes overlooked in discussions of neuroanatomy (early brain scans often left it out), but is included in our discussion because of its significance to many aspects of musical experience and performance.

As discussed in chapter 3, it is typically assumed that the auditory signal first reaches the primary auditory cortex (see Figure 4.2). The primary auditory cortex is buried in one of the brain's many folds in the temporal lobe, an area called *Heschl's gyrus*. One of the reasons why auditory research has lagged behind research in vision is because it is difficult to measure activity within this fold, as opposed to the more accessible visual cortex. Following this first stop, the auditory signal reaches many other areas, the first of which were described in chapter 3. Importantly, the stops that the signal takes after the primary auditory cortex reflect deeper aspects of listening, like the kind of information the listener is able to extract. In other words, the auditory cortex helps us know that we are 'listening to sound' and other areas (such as the superior temporal sulcus, a fold inferior to the auditory cortex) help us know what sound we are hearing.

The question of lateralization of function

In the popular press one sometimes hears about differences between the 'left brain' and the 'right brain.' Unfortunately, much of what has been communicated has been over-generalized. In technical terms, the issue has to do with *lateralization of function*, the degree to which a given behavioral characteristic (a function) is dominantly influenced by the left or right *hemisphere* of the brain. The opposite of a function that is lateralized is one that is *bilateral*, i.e., governed in equal parts by left and right hemispheres. The two hemispheres do communicate with each other, via a network of cells called the *corpus callosum* between the two hemispheres (not shown in Figures 4.1 and 4.2). However, research by Michael Gazzaniga and Robert Sperry (e.g., Sperry, Gazzaniga, & Bogen, 1969) has shown that certain cognitive functions may be dominated by one hemisphere or the other. This research was conducted on persons whose hemispheres did not communicate because the corpus callosum was severed in order to prevent the propagation of life-threatening seizures. For such persons, information received by the left hemisphere is not shared with the right hemisphere, except when both hemispheres have sensory access to the same information (for instance, if a sound is heard by both left and right ears). In this classic research, it was found that persons who had this operation could name an object that was shown to their right visual field, which transmits information to the left hemisphere, but not when it was shown to the left visual field, which transmits information to the right hemisphere. Thus, persons whose hemispheres fail to communicate can only name something when the left hemisphere has access to that information. By contrast, the right hemisphere appeared to be dominant for spatial reasoning tasks (e.g., drawing a picture of the object).

Early research in the neuroscience of music focused on whether musical pitch processing was also lateralized, and if so whether it is lateralized to the left hemisphere as is language. The implications are important for understanding whether the experience of music and language is governed by similar mechanisms. If, for instance, one were to find that music was strongly lateralized, but governed by the right rather than left hemisphere, the implication would be that music and language were independent. This has indeed been claimed, as we will discuss later in this chapter.

However, as is often the case, the story is more complex. Early evidence, in an era before PET and fMRI (techniques which will be discussed in the next section), suggested that nonmusicians process musical pitch in the right hemisphere but musicians process it in the left hemisphere (Bever & Chiarello, 1974). A similar shift in lateralization was found in an anatomical study which suggested that musicians with 'absolute pitch' (also called 'perfect pitch,' see chapter 5) have an anatomical bias for the left as opposed to the right brain in an auditory association area known as the *planum temporale* (see Figure 4.2), positioned towards the rear (posteriorly) of the temporal lobe (Schlaug, Jäncke, Huang, & Steinmetz, 1995). However, it should be noted that the data of Schlaug and colleagues are sometimes misrepresented. The leftward bias they found actually resulted from a reduced cell density in the right planum temporale of absolute pitch possessors, not from increased cell density. This leads to an important take-home message: Skill acquisition does not always increase brain size and/or activity; sometimes it actually makes one's brain more efficient and as a result reduces these characteristics of it!

Further, more recent studies do not tend to focus so much on locating specific 'music centers' in the brain, or identifying left–right hemisphere differences between musicians and nonmusicians, but on more fine-grained differences that may explain variations in brain activation patterns in music perception or performance. For instance, might the way one is first taught music, or the specific instrument(s) one learns to play, or one's style of listening to music, affect how music is processed in the brain (see for example Altenmüller, 2003)?

Neural plasticity

One reason why music training and performance is of interest to neuroscientists is because of its potential to demonstrate *neural plasticity*, the impressive capacity of our nervous system for change. In focusing on changes within the brain, neuroscientists have examined changes to *cortical maps*. Some time ago, Wilder Penfield (1891–1976) explored the brains of people who were about to have surgery for epilepsy. In order to determine what aspect of function could be damaged by this surgery, he stimulated different parts of the brain. This led to the discovery that regions of the motor cortex (posterior section of the frontal lobe) had a direct relationship to muscles in the body, and that the spatial mapping of the brain had some relationship to the spatial mapping of the body. For instance, the layout of the motor cortex's 'hand region' (its approximate location is shown in Figure 4.2) formed a hand-like shape. Furthermore, the amount of space devoted to these representations related to the degree of precision in motor control: for instance, more space was devoted to the hands than to the feet. Given these results, one might expect that musical training may influence the neural representations devoted to the hand, or some other effector used to perform music.

Indeed, the brain's 'hand region' does seem to reflect the kind of skill in hand use that one has gained through training. Enhanced representations of the fingers used to play a particular instrument seem to occur in musicians as they learn an instrument. For instance, the area responsible for controlling the left hand (which is in the right hemisphere) responds more strongly to tactile stimuli among violinists (who use that hand for fingering) than among those without such training (Elbert, Pantev, Wienbruch, Rockstroh, & Taub, 1995). In another study, it was found that anatomical asymmetries in the brain varied with both the extent and type of musical training (Bangert & Schlaug, 2006) in a way that reflected the use of left versus right hands for pianists (more right-hand representation in the brain, as most piano music requires more dexterity in the 'melody' hand than left-hand accompaniment) versus violinists (more left-hand representation, as playing violin requires more movements of the fingers of the left hand than the right hand, which holds the bow). The effects of music training and the neural mechanisms involved in executing a music performance will be discussed further in chapters 10 and 11.

Methodologies of cognitive neuroscience

We now consider various techniques that scientists use to get a glimpse into the workings of the brain. As a starting point, it is good to consider how people understood the brain prior to the twentieth century. In the ancient world, opinion was divided between those such as Hippocrates (460 BC to 370 BC) who held that all cognitive functions and all kinds of sensations and feelings originated in the brain and those such as Aristotle who credited the heart and liver with at least some of these. By the end of the eighteenth century, the dominant role of the brain as the instrument for thought, experience, and action was generally accepted. Today there is little doubt that Hippocrates got it right. In any case, we will see that our current techniques still owe a lot to historical ancestors who, while lacking our fancy modern equipment, articulated the basic logic of neuroscience and even discovered many important aspects of neural function.

The functional deficit/lesion method

Until fairly recently, this technique has dominated cognitive neuroscience in general, including the study of music. The lesion method was first developed during the heyday of the Roman Empire. Galen (129–216 AD), who served for several years as physician to the gladiator school at Pergamum, noticed that gladiators who survived after having received severe head injuries often suffered deleterious changes to their behavior and/or cognitive abilities. Such observations gave rise to studies into the link between brain *lesions* (i.e., damage) and behavioral/cognitive deficits. Here is the logic behind the method: If a psychological function is disrupted by damage to a certain part of the brain, but was present before damage, one can infer that the intact region or structure is at least a necessary component of the neural resources required for the proper performance of the function.

Based on this logic, classic neurology involves investigating changes suffered by an individual following injury, stroke, or neurosurgery necessary to preserve one's life, as in extreme cases of epilepsy, and then linking these changes postmortem to actual brain damage. Advancements in these studies led to the field now commonly known as *cognitive neuropsychology*, in which changes after brain damage are used to make inferences about the neural mechanisms underlying thought. Such an approach is often associated with the study of single cases, although as a research topic becomes more well established the proportion of group studies typically increases. Persons suffering brain damage (individuals or groups) are conceptualized as a 'treatment group' (the treatment in this case being brain damage!) and are compared to a control group comprising persons with similar backgrounds and of a similar age. If this comparison shows loss of a particular mental function (e.g., discriminating two pitches), while another mental function (e.g., discriminating two durations) is spared, there is said to be a *single dissociation* with respect to the function of the brain region that has been damaged. In the example given, there would be a single dissociation between pitch and duration discrimination. This kind of technique is popular because dissociations in the functioning of a particular brain region suggest that two mental functions may be served independently by two different brain regions. Better support yet for neural independence is the *double dissociation*, in which damage to one area is seen to disrupt ability X but not Y, whereas damage to a different area disrupts ability Y but not X. We will encounter an example of a double dissociation later in this chapter.

However, there are shortcomings to the lesion method. Lesions, whether from accidents of nature or used to guide surgery, are not inflicted in order to answer a scientific question. There are ways to perform lesion studies with greater internal validity, though each faces certain issues. First, research with nonhuman animals occasionally involves the use of lesioning. This kind of research, however, confronts obvious ethical problems. Second, a new technique known as *transcranial magnetic stimulation* (or TMS) uses a localized and intense magnetic field to create a temporary 'virtual lesion' without any damage to the brain. TMS, however, is still in its early days.

Electroencephalography

The first technique used to track brain activity (as opposed to representing just its anatomy) was the *electroencephalograph, or EEG* (as shown on the left panel of Figure 4.3a). EEG measures changes in electrical activity of the whole brain by placing electrodes on the scalp (as opposed to depth electrodes, which are inserted into the scalp and are rarely used). The transmission of impulses within neurons leads to fluctuations in the electric field surrounding neurons, as positive charged ions flow in and out of specialized gates in the neuron. The usual equipment allows for the identification of four distinct rhythmic patterns: Delta ranges up to 3 Hz; Theta from 4 to 7 Hz; Alpha from 8 to 12 Hz and Beta from 12 to 30 Hz. Each pattern is associated with certain cognitive and motor activities, and the distribution of the patterns in the total waveform of field fluctuations changes with age. Changes in these rhythms occur almost immediately as a person engages in some activity.

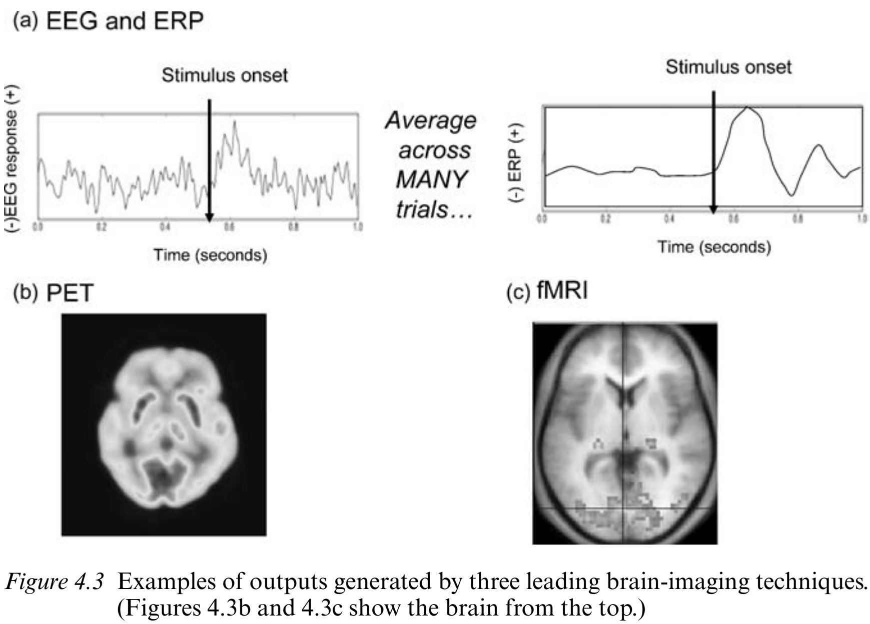


Figure 4.3 Examples of outputs generated by three leading brain-imaging techniques. (Figures 4.3b and 4.3c show the brain from the top.)

Sources:
4.3a http://en.wikipedia.org/wiki/Electroencephalograph
4.3b http://en.wikipedia.org/wiki/Positron_emission_tomography
4.3c http://en.wikipedia.org/wiki/Functional_magnetic_resonance_imaging

They also change during the onset of an epileptic fit so they have important clinical applications. Steady-state oscillations in brain activity are measured through EEG often to gauge the attentiveness of an individual. Our understanding of sleep stages, for instance, has resulted in part from measurements of the waveforms detectable in the brain through the night. The EEG plot shown in Figure 4.3a would likely result from an alert, awake individual.

In cognitive neuroscience, however, people typically do not examine steady-state EEG but instead focus on changes to the steady state that result from a particular mental process. Note that at one point the overall pattern of brain activity in Figure 4.3a shifts in the positive direction, just after a stimulus is presented to the participant (which could be a single tone that sounds out of tune). This is the kind of change to the EEG pattern that would be of interest to most neuroscientists working with musical stimuli. Such changes in activity are called *event-related potentials (or ERPs)* because a change in electrical potentials results from something a person perceives or does. As Figure 4.3a indicates, ERPs usually arise from a process of averaging across EEGs collected across many trials, leading to a smooth pattern of change in brain activity over time. The process of averaging, in theory, should eliminate all the little changes in electrical signals that reflect causes other than the stimulus, and in Figure 4.3a you can see that there is a large shift to a positive polarity that occurs, on average, shortly after the stimulus occurs. ERPs boast excellent temporal resolution and as such are used by researchers who are particularly interested in mental processes that occur rapidly (such as a 'surprise' response). On the other hand, ERPs are not good at locating the precise brain region associated with a process (i.e., their spatial resolution is poor) because brain activity is recorded at the scalp whereas the underlying activity begins at the synapse and spreads throughout the brain.

Positron emission tomography and magnetic resonance imaging

Newer brain-imaging techniques work on the premise that brain regions require more energy-rich glycogen, and more oxygen necessary to metabolize it, when they are engaged in a particular activity. These techniques result in multicolored images of the brain (though the brain is actually partly white and partly grey). There are two such techniques; the older one is called *positron emission tomography, or PET* (illustrated in Figure 4.3b).

In the PET procedure an atom of a radioactive isotope of a common element, usually fluorine, is attached to a molecule of glucose, the blood sugar that provides the fuel for brain activity. This material is injected into the blood stream of the human being under study. When a region of the brain is activated in performing some cognitive or motor task there is an uptake of glucose, and with it some of the molecules seeded with the radioactive fluorine. Radioactive fluorine atoms decay, emitting a positron (a positively charged electron). The positron soon meets a negatively charged electron. They annihilate one another emitting a gamma ray. The human subject is inside a ring of gamma ray detectors. Projecting back along the line that the gamma rays must have traversed to reach each detector, the location of the incident of mutual annihilation can be determined within a three-dimensional grid. In this way, PET offers a direct and highly sensitive measure of activity above the resting base line for each activated region of the brain. At the same time, PET images are not sharp because of the distance that a positron may have to travel before it meets an electron. Gamma rays are not emitted from the exact location at which the tagged glucose molecule was taken up,

More recently, a technique based on an entirely different physical principle has been developed known as *magnetic resonance imaging (or MRI)*. Some people have an anatomical MRI when they go to the hospital. An anatomical MRI measures tissue density in internal organs. An intense magnetic field (not thought to be harmful) is used to align the magnetic fields of individual hydrogen atoms, which are normally randomly distributed, either parallel to or anti-parallel to the field. However, a few hydrogen atoms do not line up this neat way. A radiofrequency energy pulse is directed to the body site that is being studied. The energy of this pulse alters the rate of precession of the spinning hydrogen nuclei of the 'maverick' atoms. When the pulse is switched off the hydrogen nuclei return to their original alignments, emitting their energy as a radiofrequency signal back to the machine, where it is picked up by receivers. The strength of the response signal is proportional to the density of the material being studied. By measuring the energy emitted when these perturbed hydrogen nuclei return to their original orientations, an image of tissue density can be reconstructed. This process, however, does not capture ongoing changes in the tissues being studied.

Exploiting another curious feature of the body, *functional magnetic resonance imaging (fMRI)*, illustrated in Figure 4.3c, makes possible the detection of the location of activated regions of nervous tissue such as the brain. Whereas MRI provides a high-resolution image of the brain's anatomy, fMRI shows activity related to blood flow. Specifically, when a region of the brain is active it takes up more oxygen than when in the resting state. This is reflected in a change in the proportion of oxygenated and deoxygenated hemoglobin in that region. Since deoxygenated hemoglobin weakens the radiofrequency response to an MRI intervention and oxygenated hemoglobin does not, the net result of the change of proportion of the two versions of the hemoglobin molecules is a difference in response proportional to the oxygen taken up which is, in turn, proportional to the increased activity of that region. By repeatedly scanning the region at about 2 second intervals, the relative activity of brain regions is easily mapped. This technique is known by the acronym 'the BOLD effect,' that is the 'blood oxygenation level dependent' effect. The fMRI image is blurry, whereas the MRI image is impressively detailed; therefore researchers usually use both images and superimpose the fMRI image on an MRI image.

As might be expected, most fMRI studies have employed adult participants; an example of a participant taking part in an fMRI study of music performance (and therefore playing a keyboard) is shown in Figure 4.4. However, researchers have recently begun to modify these procedures for use in research with child participants. For instance, Overy and colleagues adapted the procedures by preparing child participants with simple task instructions ('be a listening detective'), providing a cartoon story of a child having an fMRI scan, allowing children to take a soft toy into the scanner with them, shortening scanning runs, and using techniques that reduce the loud fMRI scanner noises (Overy, Norton, Cronin, Winner, & Schlaug, 2005).

Note that neither PET scanning nor fMRI directly measures brain activity, the firing of neurons, but instead measures the level of functioning of a support system for quite large numbers of neurons. PET identifies increased glucose uptake and fMRI detects increased oxygen uptake. Thus these techniques provide indirect measures of brain activity.

Figure 4.4 Experimental participant with keyboard, preparing for an fMRI experiment concerning music performance. Prior to the experimental session the participant will be moved back so that her head sits within the bore of the scanner.

Source: Photograph used with permission from the participant and the Buffalo Neuroimaging Analysis Center. Copyright © Peter Pfordresher.

Example of music and neuroscience research: The music–language link

Having given an overview of the basic brain structures and methods for studying the brain, we now turn to a specific topic as *just one example* of many intriguing research programs in neuroscience of music. One of the central questions for cognitive neuroscientists studying music (as well as the music psychology community in general) is the degree to which music and language share neural and cognitive resources. Certainly there are many human activities which share basic motor or perceptual characteristics with music, like pushing buttons and listening to the rain fall. However, researchers are drawn towards the music–language link for several reasons. First, both music and language are ways of communicating, though they communicate different kinds of information. Second, music and language both constitute richly complex sources of information that are primarily communicated via the auditory domain (though written representations of the constituent sounds have been developed in both domains). Third, and perhaps most important, whereas the importance of language to our species' survival is clear, the importance of music is a highly debated point (Cross, 2003, 2005; Huron, 2003; Pinker, 1997). That is, people are interested in whether music is *adaptive* in an evolutionary sense. If music and language are served by common brain structures, it becomes more likely that both behaviors originate from some kind of common behavioral precursor in other animals (cf. Brown, 2000).

Thus a major point of debate in research on music and language concerns the degree to which these behaviors draw on a common pool of resources, neural and/or cognitive. Evidence to date is mixed, allowing for no firm conclusion at this point. Ultimately the best account is likely to fall between the poles of complete integration and complete independence (cf. Patel, 2008). In this context, studies to date can be thought of as starting to form an idea of how much integration exists.

Evidence for integration

We will first review some studies supporting 'integration' – the idea that similar brain regions allow us to communicate with either language or music. Some of the first evidence for integration stemmed from research using ERP technology. It has been known for some time that deviations from one's expectations in language (e.g., hearing 'the pizza was too hot to cry') cause an increased negative polarity in the EEG response about 400 ms after the unexpected word ('cry'). A similar effect has been found in various studies with musical violations (e.g., hearing a pitch that goes out of key; see Besson & Schön, 2003 for a review). However, a different component responds to musical violations – positivity about 600 ms after the unexpected note. Significantly, this late positivity in language is associated with deviations that are syntactic (i.e., grammatical) rather than semantic (word meaning). This result is not surprising given that, unlike words in language, notes in music do not refer to actions and events in the outside world but the structure of music does follow rules that can be considered 'grammatical' (cf. Lerdahl & Jackendoff, 1983, discussed further in chapter 7).

What about the neural localization of music versus language? Some evidence suggests overlap here too. Various findings from studies using fMRI show brain activation in the vicinity of Broca's area (ventral part of the left frontal lobe) – long considered the locus of speech production. For instance, the presence of unexpected chords in music stimulates activity in Broca's area (Koelsch, Gunter, Cramon, Zysset, Lohmann, & Frederici, 2002). In another study, people heard normal or scrambled versions of classical music that was either familiar or unfamiliar to participants (Levitin & Menon, 2003). Listening to normal music, as opposed to scrambled variants, elicited activation in the pars orbitalis region (see Figure 4.2), a region near Broca's area also traditionally associated with language processing. These findings support a newer view of Broca's area and surrounding regions as being responsible for processing *syntax*, rules for ordering sequences in ways that are coherent to the perceiver, regardless of whether the rules obtain for musical or linguistic sequences.

Finally, there is some evidence from patients with lesions that deficits in music processing may affect the ability to process 'music-like' characteristics of language. For instance, two stroke victims who suffered lesions to the temporal lobe, including the auditory cortex, were found initially to experience profound deficits in music recognition without *aphasia* or loss of language ability (Patel, Peretz, Tramo, & Labreque, 1998). However, further investigation suggested that one of the two patients was unable to discriminate sentences based solely on intonation – for instance, in the pattern of rising and falling in pitch that can distinguish a question (rising pitch) from an answer (falling pitch). Later studies have attempted to determine whether a similar deficit is found among persons who do not have brain damage but may suffer from *congenital amusia* – an apparently inherited deficit in music perception (e.g., Peretz et al., 2002, see chapter 11). Results have been mixed; some data suggest that congenital amusics can discriminate sentences based on intonation (Ayotte, Peretz, & Hyde, 2002) whereas other data suggest a deficit. For instance, two independent studies reported by Patel and colleagues found that about 30 percent of British and French-Canadian amusics had difficulty discriminating questions (e.g., 'She plays the flute?') from statements ('She plays the flute.') based on final rising or falling pitch glide (Patel, Wong, Foxton, Lochy, & Peretz, 2008).

Evidence for independence

In contrast to the research discussed in the previous section, other evidence favors an account that treats language and music as neurally separate (e.g., Peretz & Coltheart, 2003). Studies by Perry (1993) have shown a preference for the right hemisphere, whereas the left hemisphere is typically dominant for language. He reports that 'perception of novel pitch sequences by musically experienced subjects, under conditions demanding attention to the entire melody, consistently enlists specialized processors in the non-language dominant hemisphere' (p. 102), that is the right hemisphere for right-handed people. Such evidence suggests that melody recognition involves a different area of brain activity from the recognition of syntactical structures of linguistic forms. This finding is also consistent with a new interpretation of hemispheric specialization, that the left hemisphere is good for processing fine-grained detail (e.g., what does this speech sound mean?), whereas the right hemisphere is good for processing more global information (for instance when one is attending to an entire melody). More on this distinction will be discussed in chapter 5.

An early PET study considered the way in which the brain processes speech when one attends to pitch qualities of speech, similar to when we listen to music (Zatorre, Evans, Meyer, & Gjedde, 1992). More frontal areas of the brain became active when people focused on pitch qualities of speech, as opposed to when they listened without any fixed intention ('passive' listening). Also, the activated areas tended to be more to the right side of the brain when listening to pitch qualities than when listening to phonetic qualities. In fact the activations found by Levitin and Menon (2003, who argued for integration) were predominantly left lateralized.

The aforementioned finding about attending to the pitch of syllables (Zatorre et al., 1992) brings up the question of whether people hear pitch in language using the same brain functions as they do when they hear pitch in music. Of course the fact that people hear pitch in both cases suggests that the answer might be 'yes.' However, the use of pitch in music and speech is different in many ways. Pitch varies rapidly in speech, but is often held out for longer durations in music. Furthermore, pitches are heard as representative or not representative of tonal categories in most music – for instance, we hear pitches as representative of tonic, dominant, etc. when listening to tonal music (see chapter 5). Such categories for pitch are nonexistent in many spoken languages. Or are they? In addressing this issue, researchers have focused on a class of languages that does use a framework for categorizing pitch, albeit not in the same way as in a musical scale: tone languages.

Tone languages use pitch (referred to as *lexical tone*) to convey word meaning. In Mandarin, for instance, the syllable /ma/ could mean 'mother' or 'horse' depending on the accompanying pitch. In a recent fMRI study (Wong, Parsons, Martinez, & Diehl, 2004), listeners who either spoke Mandarin (as well as English) or spoke only English heard pitch contours that could be either English or Mandarin. English monolinguals processed both contours in the right insular cortex (deep within the brain, at the juncture of the temporal, parietal, and frontal lobes, see Figure 4.1). Likewise, when bilingual Mandarin/English speakers heard the pitch contour in an

English contour, activations appeared in the right insular cortex. However, when bilingual Mandarin/English heard the pitch contour in a Mandarin context brain activations switched from the right insula to its left homologue (i.e., the same structure but on the left side). The implication here is that speakers of tone languages treat pitch as functionally different within that tone language, whereas pitch within languages like English has a more 'music-like' role and is processed accordingly. However, you need to know a tone language in order to process its pitch in the brain's 'language' area – so the distinction is not in the stimulus, but in the mental processing of that stimulus.

Finally, we note that case studies of persons with brain damage typically support the 'independence' position. A number of lesion cases (primarily stroke victims) have been documented in which brain damage hinders language ability (which, as mentioned earlier, is referred to as *aphasia*) while sparing musical abilities (summarized by Marin & Perry, 1999). A smaller number of lesion cases have been found in which brain damage hinders musical abilities while sparing language. For instance, Peretz (2003) points to the historic case of Russian composer Shebalin, who suffered a series of strokes in the left hemisphere. Although unable to speak or comprehend language, his musical skills (including composition) were left intact. In contrast, Peretz documented the case of I.R., who suffered bilateral brain damage. This resulted in lasting deficits in her perception and memory of music, although leaving I.R.'s language skills intact. This is an example of a *double dissociation*, as discussed earlier, and is typically considered to be strong evidence for the independence of two neural functions.

What is one to make of these findings? First, recall that that there is an important difference between the necessity (revealed by lesion studies) and activity (revealed by brain imaging) within brain structures while one performs a task. A certain brain structure may contribute to a cognitive task (e.g., music listening) while not being absolutely necessary to performing that task. In other words, brain structures may be specialized for a function A but may help out with function B, just like an office worker may specialize in processing purchase requests but may help with the annual reports. Patel (2008) has referred to this distinction as reflecting the difference between *representation* and *resource*. Whereas music and language may draw on a common pool of resources when a person engages in perception and/or production, the underlying representations that allow each to work may not be shared. Clearly, there is still much ground to cover and many more fascinating questions to address regarding the possible neural links between music and language.

Coda

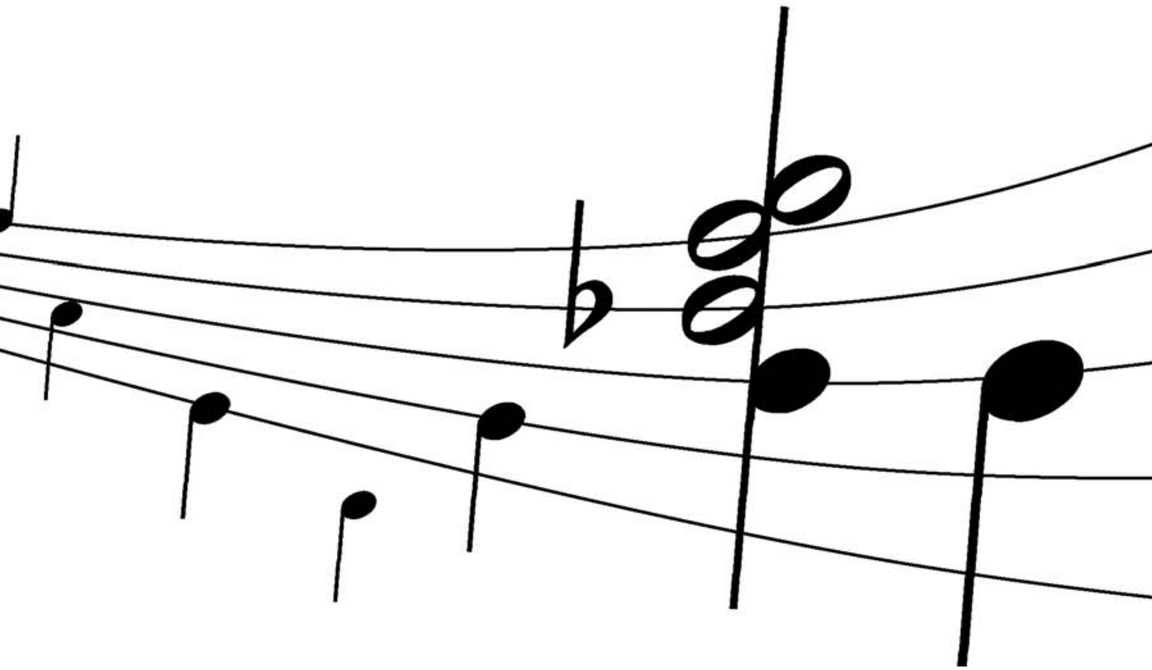
It is astonishing how much more we know about the brain now than we did only a few decades ago. It is important to temper this enthusiasm with good scientific sense. Neuroscience is well equipped to answer questions regarding the neural processes that support the perception of sound as music. It is important to realize, though, that most of the underlying theoretical issues originated in research that relied on measurements of overt behavior (responses on a questionnaire, performance on a perceptual task, etc.), not neural activity. Thus our vision of the future is not one in which neuroscience eclipses behavioral research into the conditions for musical experiences, but rather it is one in which the study of brain processes and the experiential and performance aspects of musical life complement each other.

In this context it is important to recognize how much we do *not* know from reading brain images or researching lesioned patients. What is the origin of the creative 'spark' that leads to a beautiful composition? Or a particularly moving or expressive performance? Is it neural activity? If so, what initiates the activity? These are deep issues and the answers will probably not be answered by localizing brain regions correlated with such complex and multifaceted activities as music composition or performance.

This chapter has briefly summarized those characteristics of the brain that appear to subserve musical experience, a discussion that will be further pursued in several subsequent chapters on perception of melody and rhythm, music practice and performance, and emotion in music, among others. Those who wish to learn more about cognitive neuroscience are directed toward the comprehensive textbook by Gazzaniga, Ivry, and Mangun (2002), or to the influential text by Kandel, Schwartz, and Jessell (1995). For a more in-depth discussion of neuroscience *and music* more specifically, readers are directed to an edited volume by Peretz and Zatorre (*The Cognitive Neuroscience of Music*, 2003), special issues of journals devoted to the topic (e.g., *Nature Neuroscience*, 2002; *Annals of the New York Academy of Sciences*, 2005), or useful review papers (for instance, Marin & Perry, 1999; Peretz & Zatorre, 2005).

Part II

The perception and cognition of music



5 Perception of musical pitch and melody

The following case of a 61-year-old Canadian man was documented by Peretz (1993):

> GL was referred to us in 1989 because of his persistent amusia . . . GL stated that he was totally unable to pick out familiar music and did not enjoy music any more. His musical background is that of a non-musician who was nonetheless an avid listener to popular and classical music, attending concerts and musical recitals very regularly before his illness . . . Contrasting with his relatively good linguistic abilities, recognition of tunes (without lyrics) was totally abolished. Out of 140 musical excerpts (including his national anthem) that are very familiar to everybody in Quebec, he could not identify a single one. (p. 27)

About 10 years earlier, the man had suffered an aneurysm on the right side of the brain followed a year later by a mirror aneurysm on the left side, after which he was diagnosed with severe *Wernicke's aphasia* and *amusia* (broadly defined as loss of language comprehension and music ability, respectively). He recovered most of his speech abilities but the amusia persisted. When tested on his ability to discriminate isolated pitches, GL's performance was comparable to matched controls. However, when pitches were presented in a melodic context, clear deficits were seen. For instance, GL had difficulty accurately identifying whether two melodies that were altered by one tone were the 'same' or 'different.' He was also unable to correctly identify whether a melody sounded 'complete' (ending on the tonic tone) or 'incomplete' (ending on a nontonic tone). As described above, he could not identify well-known melodies such as *O Canada*, his own national anthem! Peretz concluded that while processing of isolated pitches was intact, GL had 'lost access to tonal knowledge' (1993, p. 51),[1] which – as we shall see in this chapter – is absolutely fundamental to 'making sense' of melody.

The case of GL presents us with some intriguing questions about the perception of melody, a basic musical ability that most of us take for granted. How do we learn and retain the thousands of tunes we 'know'? Melodies are simply different arrangements of discrete pitches with distinctive rhythmic patterns. How are we able to remember each particular arrangement? If melodies are made up of separate pitches, what binds a melody together so that it is heard as an organized and coherent tune? These questions serve as the starting point for our discussion on melody and pitch. We will review the classic work on the perception of Gestalt wholes, explore some findings on memory for pitch and melody, and conclude with related studies in the neuroscientific research.

The ingredients of melody

In keeping with standard definitions of melody (e.g., Radocy & Boyle, 2003), we consider the term 'melody' to characterize one aspect of musical experience: the experience of a sequence of pitches as belonging together. This definition captures the idea that melodies are perceived, not in terms of their separate constituent tones, but as coherent units. While each tone of a melody reaches the listeners' ears as if it were a single bead, listeners 'thread' the beads together into continuous strands. We will turn to the topic of the perceived unity and coherence of a melody line in our discussion on Gestalt principles of perception. But first we will explore some essential elements of melody: pitch, interval, contour, harmony, and key.

Pitch

We already considered the concept of pitch at a psychoacoustic level in chapters 2 and 3. However, as we have seen in the case of GL, when pitches are perceived within the context of a musical melody our perception of pitch becomes more complex. It has been pointed out that musical pitch is a *multidimensional* percept; that is, there are multiple continua along which pitches may be distinguished (Shepard, 1982). We focus on two here. One dimension, true of pitch either in or out of a musical context, is *pitch height*, related to the frequency of vibration. However, a more musically important dimension of pitch is its so-called *chroma*. Chroma refers to the category (or 'class') represented by a certain pitch. The names we give notes in Western tonal music (e.g., C, D, E) refer to pitch chromas. Chromas are identical when separated by an octave (a 2:1 ratio of a tone's fundamental frequency) and thus constitute a distinct dimension from pitch height. The fact that tones separated by an octave inhabit the same chroma category is called *octave equivalence*. Within an octave, however, changes in chroma are also changes in pitch height; the dimensions are therefore distinct but not independent. Evidence that the auditory cortex maintains distinct representations for pitch chroma and pitch height has been reported in an fMRI study (Warren, Uppenkamp, Patterson, & Griffiths, 2003).

So far we have established that one way to characterize melodies is as a sequence of tones, with each separate tone containing particular information about pitch height and pitch chroma. However, this is probably not the way most people conceptualize melodies. Consider the fact that most individuals, even from an early age (i.e., infancy, as discussed in chapter 8), perceive a melody to be the same when it is played in a different key (a 'transposition'). If 'Over the Rainbow' is transposed from the key of C to D major, the opening phrase changes from C-C′-B-G-A-B-C′ to D-D′-C♯-A-B-C♯-D′, which changes every pitch in the sequence. Yet most people have no trouble recognizing these melodies as the 'same.' In transposed melodies, even though specific pitches may be changed, the *relationships (i.e., intervals)* between the pitches are retained, and it is these relationships that people primarily use when recognizing a melody. We now turn to ways in which the relationships between pitches may be characterized.

Interval

Intervals denote the transition from one pitch to the next. In musical practice, one more commonly hears of scale step intervals (e.g., major third, perfect fifth), which assume a tonal structure. The interval level of description is an abstraction because it loses information about specific pitches. Its significance to the cognitive representation of music is that people can typically recognize and reproduce a melody in many different keys. This accomplishment relies on the fact that people represent melodic structure in a way that is independent of pitch class. This is what is meant by the term *relative pitch* perception, the ability to recognize and remember pitches by virtue of their relationships to one another (i.e., intervals between pitches) as opposed to the specific identity of individual pitches. Individuals who possess *absolute pitch* (or AP, commonly referred to as *'perfect pitch'* among musicians) can identify the pitch chroma of isolated tones (passive AP) or produce a chroma (active AP), without being given a reference pitch. Thus AP possessors do not rely on intervallic relationships to recognize pitches. We shall return to the fascinating topic of absolute pitch later in this chapter.

Melodic contour

Melodic contour refers to the shape of a melody line, depending on whether successive pitches are rising, falling, or unchanging in pitch. Like the distinctive skyline of a city, it is the shape or outline that gives a melody its distinguishing character. Researchers are interested in melodic contour because changes in direction constitute salient points in a melody (e.g., Jones, 1987). Moreover, a performance error that results in a change to the melody's contour is far more noticeable (and far less common) than an error that does not cause such a change. GL was unusual as he had difficulty discriminating between short melodies even if the alteration modified the direction of the pitch, changing the overall contour (Peretz, 1993). However, as we will see in chapter 8, even infants can discriminate between two melodies that differ only in one pitch if it alters the melodic contour (e.g., Trehub, Thorpe, & Morrongiello, 1985).

Harmony

Harmony and melody are often referred to as if they were independent, so it is perhaps odd to see harmony listed as an ingredient of melody. However, the harmonic scheme can influence the perception of melody in many ways. For example, the harmonic structure can imply segmentation of the melody into sections by virtue of harmonic changes. When a phrase ends with an authentic cadence (harmonic progression from dominant to tonic, or chord V to I), the end of a section is strongly implied, even if the music continues after that cadence. These boundaries form a higher level of structure than do changes in interval and contour. Melody can also imply harmony even in the absence of any accompaniment. A melodic sequence of notes such as [G A G E C] gives the strong implication that if one were to play a chord along with these notes that the chord ought to be a C major tonic triad. Indeed, our ability to recognize melodies depends in part on our sense of implied harmony (Cuddy, Cohen, & Mewhort, 1981), especially for musically trained listeners (Schubert & Stevens, 2006). We have assumed diatonic scales (e.g., major, minor) in our discussion, but there are many other scale and tonal systems around the world (see chapter 15).

Key

Key refers to the kind of scale structure that a melody implies. In the diatonic scale system it can be *tonal* or *atonal*. Tonal melodies are those for which the constituent pitches imply a specific key, whereas for atonal melodies there is no such implication. If tonal, the key can be major or minor, or another mode. Key may constitute the most abstract way to describe a melody, in that a melody can often be referred to with respect to just a single key (although key modulations are common, which relegates key to a more local level). As a description of melodic structure, however, key is highly limited. There is no way to recover the structure of a melody if all one stores in memory is the key. Key places important limitations on notes that sound as if they ‘belong’ to a melody. As with contour, errors that violate key are highly salient. Similarly, out-of-key notes that are consciously added to a composition can create moments of high tension. GL was a rare exception; when asked to select a tone to complete a five-tone sequence, he preferred tones that were not part of the scale to diatonic tones.

Not all pitches are ‘equal’ within a scale. For instance, the *tonic* and *dominant* (first and fifth degrees of any scale) are structurally important, and usually give a sense of a stable center and closure (tonic) or repose (dominant). Indeed, music theorists view every tone within the diatonic scale as having a function in relationship to the other tones (as indicated by their technical names such as ‘tonic,’ ‘dominant,’ ‘leading tone,’ or ‘leading note’). Harmonically, the *tonic*, *dominant*, and *subdominant chords* (chords I, V, and IV) in any key are structurally most important. Many simple songs can be harmonized simply by using just these three chords. Simply through exposure to music, even listeners with absolutely no musical training possess some general knowledge structures of tonal music that are activated when listening to music, enabling them to form basic expectations about where the next note or phrase might go, or if a song has come to a 'good finish.' GL represents a very rare illustration of how, in the absence of tonal knowledge providing the context, melodies lack meaning. While explanations for the structural importance of some tones of the scale over others are provided by music theorists and physicists, we will focus on a study exploring the perceptual bases in our discussion of 'tonal schemata' and Krumhansl's (1990) research later in this chapter.

The perceptual organization of melodies

Earlier, we compared the separate tones of music to loose beads, which listeners thread together into strands. When listening to more complex music beyond the single melody line, a more complicated network of these strands of beads emerges. In many cases, there may be many different ways that the beads may be organized, as a complex work is open to many interpretations. When listening to a pop song, jazz standard, or symphonic orchestral work, there is a dynamic interplay of melody and harmony lines, and relationships between voices and instrumental parts. In part, it is this ability to organize incoming musical input into coherent units that makes music listening such a rich and enjoyable experience. In this section, we take a closer look at how it is that listeners organize discrete musical sounds into coherent wholes as a composition unfolds, focusing on the perception of melody.

Gestalt principles of perception

The word *Gestalt* is German for 'whole form.' The study of the conditions under which 'gestalts' are perceived in a sequence of elementary parts is the main aspect of Gestalt psychology that is of interest to us. In other words, gestalts are the organized structures that emerge from the physical stimuli in our environment.

This development was initiated by Max Wertheimer in Berlin before the First World War and extended by others, notably Kurt Koffka (1896–1941) and Wölfgang Köhler (1887–1967). The Gestalt pioneers began systematic studies of how the elements in an aggregate must be arranged in order for it to display an emergent property, to be seen or heard as a whole. According to the Gestaltists, the perception of the emergent or whole property of an aggregate and of the elements contributing to that whole influence each other. We see patterns in groups of elements, and the perceived groups in turn influence our perception of the elements (Wertheimer, 1938).

The following are examples of basic Gestalt principles describing how elements group to form an aggregate. These principles, originally designed to be applied to vision, follow from a single superordinate principle called *Prägnanz*, which dictates that perceptual organization conforms to a form that is simple and symmetric (Koffka, 1935, p. 110):

> *Proximity*: Other things being equal, the elements that are near to one another tend to be seen as a group.
>
> *Similarity*: When more than one kind of element is present those which are similar tend to be seen as a group.
>
> *Closure*: When a pattern is incomplete, there is a tendency to perceive it as whole and complete by 'closing' the gap.
>
> *Good continuation*: Smooth continuity is preferred over abrupt changes of direction.

Although first applied to visual phenomena (as illustrated in Figure 5.1), these principles have a more or less direct application to the perception of musical wholes, the components of which are tones.

Gestalt principles and music

In its application to music, the *Principle of Proximity* may explain how a series of tones is perceived as a melodic line, as opposed to a series of

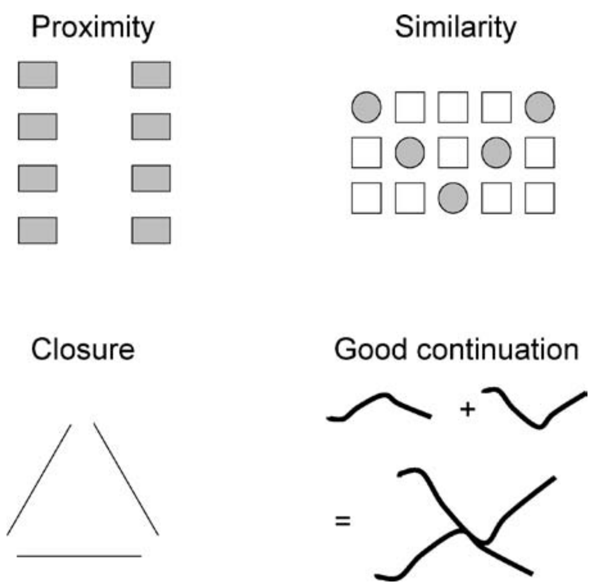


Figure 5.1 Examples of Gestalt principles, with visual examples. Copyright © Peter Pfordresher.

unrelated and disconnected tones. In music, 'nearness' may be applied in several forms. Tones that are close together in *pitch*, *time*, or *space* tend to be perceived as a group. For instance, tones separated by large leaps in interval, separated by rests or pauses, or originating from different spatial locations (e.g., music coming from left versus right loudspeakers) tend to serve as boundaries between groups. Of these, proximity in pitch appears to be one of the most important factors in governing grouping of auditory streams. Melodies in both Western and non-Western music tend to be comprised of small intervals and relatively narrow pitch range (e.g., see Dowling, 1968).

Indeed, Diana Deutsch has demonstrated how 'octave-scrambled' presentations of familiar melodies are difficult to recognize. Although most effectively demonstrated in Deutsch's CD *Musical Illusions and Paradoxes* (Deutsch, 1995, Track 19–23 'mysterious melody'), the effect may be easily demonstrated on a piano keyboard by playing a well-known melody on a piano in an 'octave-scrambled' fashion. Simply preserve the original notes and rhythms but play each note in a different octave (most effective if using full range of the keyboard, and selecting octaves without any pattern). Not only will most listeners be unable to identify the 'octave-scrambled' melody, but most will also not experience the sequence as a coherent 'melody.' Due to the gross violation of the Principle of Proximity, the tones do not cohere into a melody but seem to 'fly apart'!

The same principle is at play in a classic study by W. Jay Dowling (1973). Dowling investigated the conditions under which listeners are able to separate two 'interleaved' (or intertwined) melodies. Interleaving was accomplished by playing alternate tones between two familiar melodies (i.e., first tone of melody A, first tone of melody B, second tone of melody A, second tone of melody B, and so on). This musical device is often used by composers and orchestrators. Here is a visual analogy of interleaving, employing interwoven sequences of letters.

INTERLEAVED

TWOMELODIES

When interleaved, the result would be:

I T N W T O E M R E L L E O A D V I E E D S

Dowling showed that when two familiar melodies are interleaved, they are difficult to distinguish when the melodies presented are overlapping in pitch ranges. When played in different pitch ranges, however, the two groups of tones are heard as two distinct melody lines. Many examples of this effect can be found in musical works, in which alternating between high and low registers may create the impression of two melodic lines played by a single instrument (*virtual polyphony*) as exemplified by Bach's solo violin works (e.g., Davis, 2006), or in single voice as in yodeling. Due to our tendency to group pitches by proximity, two melodic lines are heard, as opposed to a single jagged melody alternating between high and low tones.

The *Principle of Similarity* also plays a significant role in the perception of music. There are times when we attend to a single voice or instrument. But for the most part, whether listening to a local garage band or a symphonic orchestra, we organize the input of individual sounds into organized groups such as ‘melody’ and ‘accompaniment,’ ‘strings,’ ‘winds,’ ‘brass,’ and so on. In such circumstances, it is often timbre that serves as the point of similarity that allows the listener to segment sounds into different instrument groups.

Another application of the Principle of Similarity in music is sequential similarity, by which repetition and variation of themes, motifs, and relationships among musical ideas facilitate our perception of a musical work as a coherent whole. The elements in a visual array are often simultaneously present for inspection of repetition or patterns, but many forms of musical similarity unfold sequentially, in time. Ravel’s *Bolero* is a marvelous study in repetition, with its insistent theme repeated throughout the piece over an ostinato rhythm. Sometimes, as in the child’s round ‘Row Row Row Your Boat’ or Eric Clapton’s ‘Layla’ (which opens with six exact repetitions of a riff) the repetition is quite apparent to a perceptive listener. In other cases, such as the doubling of the theme in other keys as the orchestra augments in *Bolero*, or appearances of the subject in a complex three- or four-part Bach fugue, the relationships may be more difficult for the listener to perceive. It is our ability to hear repetition, patterns, and other musical relationships that lends a sense of unity and order to a work of music as a whole.

It should be noted, however, that these principles of perceptual coherence may hold for short simple melodies but perhaps not for more extended works (see Levinson, 1997). For instance, when Tan and Spackman (2005) created ‘patchwork’ compositions by linking extracts of music by different composers (e.g., Schumann–Liszt–Chopin) together with abrupt changes in pitch, key, harmony, tempo, style and little repetition of melodic ideas, few (musically trained or untrained) listeners noticed that the music had been structurally altered. In another study, Tan and her colleagues found that after four repeated hearings, listeners even came to prefer the ‘patchwork’ compositions to intact compositions, and rated them higher in musical unity (Tan, Spackman, & Peaslee, 2007). In a study in which ‘hybrid’ compositions were created from combining pieces of different Mozart piano sonatas, Eitan and Granot (2008) also found that with repeated exposure, listeners showed a preference for the ‘hybrids’ and, surprisingly, musically trained participants preferred the hybrids more than untrained participants. It is important to keep in mind that Gestalt principles that may easily be demonstrated with short melodic sequences may not always apply to more complex and extensive musical works.

The *Principle of Closure* is exemplified in the way that finality as completion of a melody is expressed by resolution to the tonic of the scale, either in the melodic line or in the actual or implied harmonic progression. We are reminded of the almost surely apocryphal story of how a visitor got Mendelssohn out of the bath. As the story goes, a visitor was informed that Herr Mendelssohn was in the bath and could not see him. The guest went to the piano and repeatedly played the first seven notes of the scale of C major. After several repetitions, Mendelssohn suddenly appeared in his bath robe – and completed the last note of the scale![2] The leading tone (or seventh note of the diatonic scale) sets up such a strong expectation to resolve to the tonic that most listeners imagine the last note, 'closing' the scale with a return to the tonic – much like one mentally fills the gap to close an incomplete circle. In Mendelssohn's case, however, it was not enough to mentally complete the scale. According to the story, he was absolutely compelled to *play* that tonic note! Such strong expectations based on the functions of the tones in a scale were mentioned earlier; pitches in a melody are associated not only with a current value but with a kind of musical velocity. In the case of GL, who did not seem to hear the pitches in a tonal context, the principle of closure was not at play; he did not seem to feel the 'pull' to the tonic.

Experimental evidence for closure in music was reported by DeWitt and Samuel (1990). The authors tested whether an effect similar to phonemic restoration in language occurs in music. *Phonemic restoration* (Warren, 1970) is a powerful effect in which listeners continue to hear intact speech even when a phoneme is replaced by noise. Rather than hearing the noise where it occurs, the listener experiences the speech sound as if continuing *through* the noise. This effect is analogous to 'filling in a gap' visually, and tends to be stronger at the end of sentences than at the beginning (suggesting that the process of filling in requires knowledge of the context, which is referred to as 'top-down processing'). In their investigation of *perceptual restoration* in music, DeWitt and Samuel (1990,[3] expt. 4) created 10-tone scales (octave scale plus next two tones) with one tone either present but accompanied by noise or replaced by noise. Listeners had greater difficulty discriminating whether the tone was present or absent with the noise if the distorted tone was the expected pitch that would complete the scale (as opposed to a random pitch). They were also more likely to 'restore' a missing pitch as the number of tones of the scale before the distorted tone increased. Thus the findings showed 'an increase in restoration associated with increased musical expectations' (p. 141).

While Gestalt principles account for our natural tendency to group musical sounds in organized ways, they also explain the principles by which particular auditory streams of musical sounds are heard as distinct from the rest. In particular, this pertains to the question of how a melody line (the 'figure') can be heard against sometimes very complex and busy accompaniments ('ground'). For a thorough discussion of auditory stream segregation in instrumental and vocal ensembles, the reader is referred to Albert Bregman's *Auditory Scene Analysis* (1990), especially chapter 5 and pages 490–502.

In the first chapter to this important book, Bregman states that 'the Gestalt principles of grouping can be interpreted as rules for [auditory] scene analysis' (p. 24).

The scale illusion

In most cases, our natural tendency to sort incoming sounds in organized ways helps us to make sense of complex auditory stimuli. Under certain conditions, however, the Gestalt principles may be in play where in fact they should not apply and this can lead to false impressions about what we are actually hearing. One of several interesting auditory illusions described by Diana Deutsch (1975), and used here as an illustration, is the *scale illusion*. In Deutsch's study, participants listened to a musical sequence through stereophonic headphones. The sequence was based on the C major scale, with the notes of the ascending and descending scale alternated simultaneously between the left and right ears, producing two rather jagged melodic lines played to each ear (as illustrated schematically in the top part of Figure 5.2).

Participants were asked to give a verbal report of what they heard through the headphones, and to sing what they heard in each ear. Not one of the 70 participants reported hearing the musical pattern that was actually played! As illustrated in Figure 5.2, most (right-handed) participants 'heard' smooth ascending and descending contours of the scale, with the higher tones in the right ear. The most common percept was half a descending C major scale which then ascended back to the C in the right ear, and half an ascending C major scale in the left ear which then descended back to the C. Thus, the

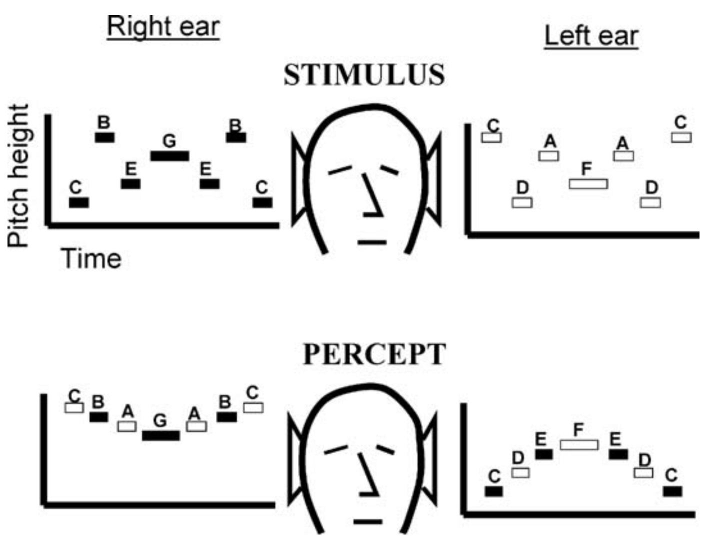


Figure 5.2 An illustration of the scale illusion. Rectangles denote tones varying in pitch and timing. Filled rectangles denote tones presented to the right ear (through headphones) and open rectangles denote tones presented to the left ear. Adapted from Deutsch (1999).

tones not only 'reorganized' into smooth melodic contours, but seemed to 'migrate' from one ear to another to do so. When (right-handed) participants were asked to reverse their stereophonic headphones so that the input formerly presented to the left ear was presented to the right ear and vice versa, most participants' reports did not change. Again, they heard a smooth descending-ascending scale pattern in the right ear, and the mirrored contour (all the lower tones) in the left ear. The responses of left-handed participants showed more variation, and no clear pattern was reported for them.

The 'scale illusion' illustrates several Gestalt principles of grouping, especially the *Principles of Proximity and Similarity*. If the tones are played in sufficiently different pitch ranges, sufficiently different timbres, or if the tone sequence does not imply an orderly scale, the effect is likely to be weakened or absent, as the perceptual principles do not act singly but often act 'in concert' or in competition with one another. For instance, it is interesting that most listeners do not hear a full descending scale in one ear and a full ascending scale in the other, analogous to our visual illustration of the *Principle of Good Continuation* in Figure 5.1. Other predispositions, such as the strength of the proximity principle and our tendency to more accurately localize high sounds in the right ear and low sounds in the left ear, may override the tendency to sort the tones into continuous lines that cross over each other.[4]

Memory for pitch and melody

Having explored the immediate perception of melody, we now turn our attention to memory for melodic information. In this section we discuss memory for pitch, contour, tonality, and an example of implicit memory for music.

Memory for pitch

It should be noted that the aforementioned studies assume that melodies are remembered without reference to specific pitches, as described earlier. This limitation makes sense given the common assumption that few people encode melodies using absolute pitch. The ability to label individual pitches that one hears without being provided with a reference pitch may be fairly rare, possibly occurring in only 1 of 10,000 persons (Takeuchi & Hulse, 1993). Perhaps in part because of its rarity, *absolute pitch* (or *perfect pitch*) is a topic of widespread fascination, in spite of the fact that relative pitch may be a more useful ability for music perception, due to the relative nature of melody. For instance, musicians with relative pitch outperformed those with absolute pitch when asked to judge whether two transposed melodies preserved intervallic relationships or not (Miyazaki, 2004). In comparison to absolute pitch, the ability to recognize pitch intervals is more common and it is widely assumed that such a *relative pitch* representation dominates the processing of musical pitch for most persons.

Recent evidence, however, suggests that most people in the general population may store some melodies, not just as abstract relational representations, but as absolute pitch representations in long-term memory. An early study demonstrating this was run by Levitin (1994) who simply asked participants to sing their favorite tune from memory, beginning wherever they wish. Surprisingly, these participants (who were not trained singers and did not possess absolute pitch) produced melodies within 1 or 2 semitones of the original key of the melody most of the time. Importantly, the songs used in Levitin's study were recorded pop tunes that listeners had always heard in the same key. This study showed that although correct *pitch-labeling* (C, D, E♭, etc.) without a reference pitch may be rare, good *pitch memory* seems to be widespread. Similarly, Bergeson and Trehub (2002) found that mothers singing songs to infants were very consistent in pitch and tempo when recordings of singing (separated by intervals of one week or more) were compared.

A plausible explanation – according to one view – would be that participants transpose melodies into the key most comfortable to their singing range, resulting in a fairly uniform distribution of produced keys. However, other studies showed that evidence for accurate pitch memory is not limited to production. For instance, Schellenberg and Trehub (2003) played orchestral theme songs from popular television programs (including *Friends*, *E.R.*, and *The Simpsons*) either in their original pitches, or shifted up or down by only one or two semitones. Most people were able to identify the original key of familiar television theme songs despite the variations being very close in pitch, leading the authors to conclude that 'good pitch memory is widespread' (p. 262). Other studies have shown that people can identify tunes from MIDI (Musical Instrument Digital Interface) recordings within one phrase (Dalla Bella, Peretz, & Aronoff, 2003), and can identify audio recordings within 200 milliseconds (Schellenberg, Iverson, & McCinnon, 1999).

Does pitch perception change with age? One of your authors (ST) was identified as having AP at around age 7, while being tested by an examiner during an aural exam for the Associated Board of the Royal Schools of Music. She was quick at identifying single pitches and could even name pitches comprising three- or four-note chords with high accuracy. During her mid-thirties, she began to make more errors when identifying pitches (especially in the extreme high and low ranges), and tended to be a semitone sharp when incorrect (e.g., hearing the lowest E♭ on a grand piano as an E). She has since discovered that this 'shift' is common. Athos and colleagues (2007) tested almost 1000 AP possessors between the ages of 8 and 70, and found that 'pitch errors [among individuals with AP] increase with age, and they tend to be sharp' (p. 14796). In fact, one 44-year-old participant did not identify a single pitch correctly during the test but was consistently one semitone sharp, a tendency he had observed since age 22. The reason for this shift is not yet known but Athos and colleagues (p. 14797) surmise that it may correspond with some 'age-dependent physiological change that alters the mechanical properties of the cochlea' (whose critical role in hearing was discussed in chapter 3). Indeed, the authors hypothesize that 'such a gradual perceptual shift is common to most people as they age, yet they are unaware of it unless they have AP' (p. 14797).

Memory for melodic contours and intervals

We mentioned earlier that melodic contour constitutes the most basic and possibly the most salient attribute of melody. As it turns out, the influence of melodic contour on memory is strong but it is also qualified. It appears that the only kind of melody for which melodic contour entirely determines memory confusions may be an *atonal* melody. When melodies are *tonal*, key also contributes to melody recognition. Dowling (1978) reported a study in which people heard pairs of melodies and had to determine whether the second melody constituted a transposed version of the first melody or was a different melody. Whereas musically trained listeners were good at ruling out an atonal melody that preserved the original melody's contour, they were not as good at ruling out 'tonal answers,' which preserved contour and key but differed with respect to constituent intervals. Tonal answers are akin to repeating a melodic theme in different registers of a common key. For instance, if one plays [C E D F G] followed by [F A G B C] the two excerpts sound quite similar by virtue of their common contour and their similar reference to the key of C, even though the last two intervals differ. Unskilled listeners are more likely to be fooled by atonal melodies that match in contour, though they too are better at ruling these melodies out than tonal answers.

The results summarized in the previous paragraph apply to melodies held in memory over the short term. As we become increasingly familiar with melodies, our memory starts to store interval information as well, and as a result people get better at ruling out different kinds of alterations. It is highly unlikely, for instance, that one would be fooled by a 'tonal answer' to 'Happy Birthday.' This finding was originally reported by Dowling and Bartlett (1981; see also Dowling, Kwak & Andrews, 1995).

Tonal schemata

Musical memory need not be specific to a particular melody – it can work in more general ways. One uncanny aspect of music perception that we often take for granted is the fact that a 'wrong note' can really stand out perceptually, even when the wrong note is very close, physically speaking (for example, in spatial location on a keyboard), to the 'correct' note. And 'nearby' wrong notes (C♯ instead of C when the key is C major) can be more noticeable than some 'distant' wrong notes (such as playing a G instead of a C when the key is C major). A broader implication of this fact is that tonal musical contexts cause listeners to categorize tones in ways that often contradict basic psychoacoustic aspects of pitch – perceptual and physical 'closeness' are not necessarily the same thing.

The implication of this research is that listeners maintain internalized rule structures (commonly referred to as *schemata*) in long-term memory and use these rule structures to interpret incoming sounds such as melodic sequences. Tonal schemata differ from those discussed earlier for melody in that tonal schemata are typically not thought to rely on or to use sequential information. Rather, tonal schemata are used to interpret the role that constituent pitches play within a melody, with respect to the key that the melody evokes. (Examples were given in our discussion of key, earlier in this chapter.) Though individual differences have been documented with respect to the way in which listeners acquire and use tonal schemata (see Smith, 1997 for a review) recent research suggests that all listeners may gain these kinds of rule systems through exposure, even if they are not conscious of doing so (Tillmann, Bharucha, & Bigand, 2000).

The leading figure in this area of research is Carol Krumhansl, whose work is extensively discussed in the book *Cognitive Foundations of Musical Pitch* (Krumhansl, 1990). Krumhansl introduced the *probe tone technique*. She presented listeners with a short tonal context, which could be a scale (major or minor), a chord, a series of chords, etc. Following this context, listeners would form some kind of categorical judgment on a tone or pair of tones that followed the context (such as rating the similarity between two pitches). Generally speaking, listeners' categorical judgments were most strongly influenced by the status of tones within the *pitch hierarchy* that was established by the context. In a tonal context, the *tonic* (this is the pitch that gives the key its name, e.g. C in the key of C major) is considered to be the 'most stable' tone, followed by the *dominant* (which is G within the key of C major) and the *mediant* (E in the key of C major), other within-key notes (D, F, A, B), and finally notes that are out of the key.

A central finding was one in which listeners categorized the similarity of two successive tones that followed the context. The rating results were displayed using a technique known as *multidimensional scaling*. The idea behind multidimensional scaling is essentially similar to that of a scatterplot. A scatterplot shows the relationship between two variables, X and Y, in Cartesian coordinates (a rectangular grid). The coordinate system is assumed and, within it, a relationship may look 'good' (e.g., suggesting a clearly linear relationship) or 'messy.' By contrast, the coordinate system in multidimensional scaling is flexible and can be altered to characterize the relationship between (or among) variables in an ideal way. Moreover, when analyzing similarity ratings every single pair constitutes a possible dimension (or variable), and thus multidimensional scaling offers a way to simplify the display of results. Thus, similarity ratings may be plotted in two or three dimensions, on a curved or flat surface, and so on. Ultimately the data are displayed in a way that forms some kind of geometric shape, suggesting that the cognitive representation of the relations between pitches are best captured by this shape.

The similarity ratings that Krumhansl collected best conformed to a conical shape, shown in Figure 5.3 (described by the author as a 'slightly idealized'

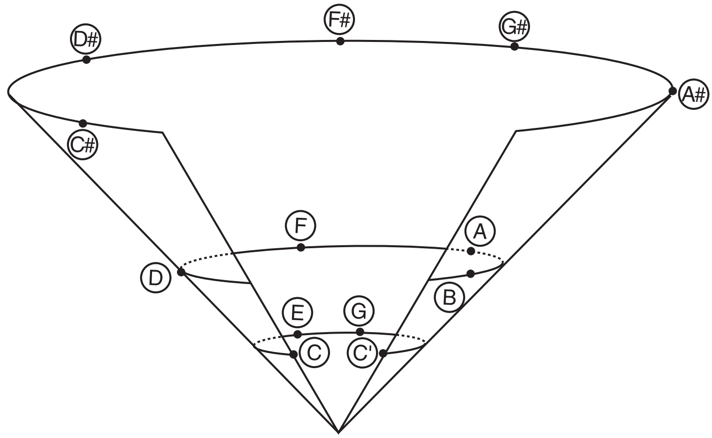


Figure 5.3 Conical representation of similarity ratings for pitches within a C-major context (Krumhansl, 1979, Figure 3).

Source: Reprinted with permission from Elsevier.

scaling solution). Though this kind of representation may look unfamiliar, what it does is fairly simple: similarities are shown as geometric distances. Note that different points are associated with pitch classes within the C-major scale (though in principle this shape could apply to any major key). Closer points are heard as more similar to each other. Thus, when the tonal context is followed by the pitch pair [C G], a participant would be likely to rate the pair as highly similar, whereas the pair [C D♯] would be rated as highly dissimilar. Note that Krumhansl's results differ from basic psychoacoustics, in that the fundamental frequencies of C and D♯ are closer (when presented in the same octave, as Krumhansl did) to each other than are the frequencies of C and G. Some have suggested that tonal structures like these do follow from a different principle from psychoacoustics, that is the avoidance of dissonances (e.g., Parncutt, 1989). For instance, minor seconds when sounded together lead to high dissonance due to partials that fall within the same critical bandwidth (see chapter 3). On the other hand, a perfect fifth (closer in the geometry of Figure 5.3) is highly consonant.

Two aspects of the geometry are of central importance. First, note that different levels of the pitch hierarchy line up at different vertical levels. The bottom level includes pitches that make up a C-major triad, which is the 'root chord' for C major. The next level higher includes pitches that are in the key of C major but are not part of the triad. These two levels are closer to each other than the second level is to the third level, which includes pitches that do not fit within C major. A second aspect of the geometry concerns the diameter of the semi-circle formed by pitch relationships at each vertical level. The bottom level (the C-major triad) forms a semicircle with a small diameter. This means that all the pitches at this level are heard as fairly similar to each other. In general, you can see that all the within-key pitches were heard as being fairly similar. By contrast, note that the cone widens noticeably at the top level. Pitches outside the tonal context were thus heard as being dissimilar not only to in-key pitches but also dissimilar to each other. The downward orientation of the cone in addition suggests that there is a kind of 'gravitational pull' towards the tonic. With respect to the way we experience music, this model has implications for notes that cause a listener to feel 'tension' or 'relaxation.' Notes near the bottom of the cone sound like endings and thus lead to relaxation, whereas notes near the top suggest that all is not finished, leading the listener to feel a tense kind of suspense.

Note that the schema laid out in Figure 5.3 is true for pitches within a single key. What about relationships among keys? A common music theory perspective is that all the keys within a single mode (major or minor) are related to each other via the *circle of fifths*, which is shown in Figure 5.4. Though the figure shows the circle for major keys, the same layout would hold for minor keys. As with the Krumhansl cone, keys that are closer together around the circle are thought to be more similar. Similarity here, in contrast to pitch, is determined by the set of pitches that make up a key. Keys that are adjacent around the circle have all pitches except one in common; for instance C and G major differ by only one pitch (F versus F♯). By contrast, C and F♯ major keys only have one pitch in common, the rest differ (only B is held in common). Thus, although the tonic pitches for C and D♯ are quite close, the keys associated with these tonics are quite distinct.[5]

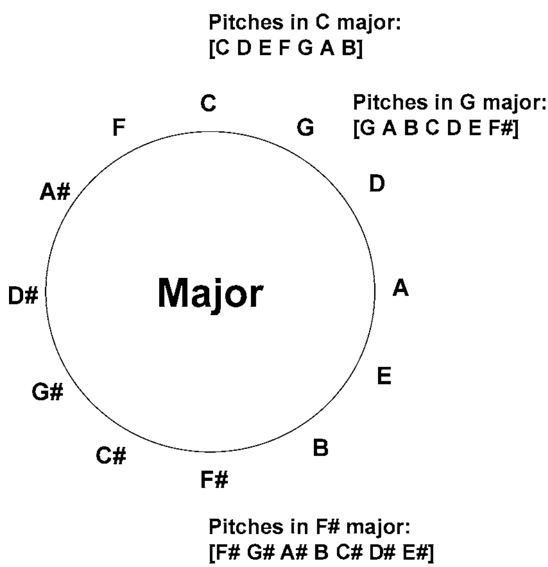


Figure 5.4 The circle of fifths for major scales. Constituent pitches for three selected keys (C, G, and F♯) are given in brackets.

Implicit memory and priming

Memory need not always involve a conscious recollection of past events (explicit memory). This can be illustrated in the role of *priming* in music cognition. The idea is that the perception of a stimulus leads you to access related items from memory. As a result, the perceptual system can more easily encode these related items for a brief period of time. Priming depends on a process psychologists call *implicit memory*. Implicit memories influence how we perform tasks but are not typically accessible to consciousness. In fact it is often difficult to verbalize an implicit memory. For instance we know how to tie our shoes based on memory, but most people find it very difficult to describe how this task is performed!

Many parallels have been drawn between priming effects in language and in music (e.g., Tillmann & Bigand, 2002, for a review). In *semantic priming* in language, the recognition of a word can be influenced by relevant material presented earlier. For example if someone is shown the word ‘donkey’ the speed with which they can distinguish ‘horse’ from ‘hoarse’ is increased (McNamara, 2005). Similarly, in *harmonic priming* in music, processing of a chord tends to be faster and more accurate when preceded by a chord that is harmonically related or ‘schematically probable,’ compared to one that is not (Justus & Bharucha, 2001; Tillmann & Bigand, 2001). This effect has been demonstrated in studies measuring behavioral responses as well as experiments employing scanning with functional magnetic resonance imaging (fMRI) (Tillmann, Janata, & Bharucha, 2003).

Repetition priming, another kind of priming simply defined as ‘a processing benefit for previously encountered stimuli’ (p. 693), has been demonstrated with melody (Hutchins & Palmer, 2008). When participants were asked to sing the last pitch of a five-tone melody, they responded faster if the pitch was included as one of the first four tones than if it was a new pitch. Response time was faster when the prime and target were closer together in time, and faster for tonic endings than nontonic endings.

Neural bases of pitch and melody perception

In this final section, we discuss some examples of neuroscientific studies on the perception and memory of musical pitch, key, and melodic contour. This approach is described in chapter 4, and Figure 4.2 may be useful in locating brain regions referred to in the following discussion.

Chroma

One issue that has intrigued researchers in music cognition is the degree to which the brain possesses fixed chroma-specific pitch categories; that is, *absolute pitch*. The dominant view, as discussed earlier, is that most people do not remember ‘absolute’ pitch, that is, they do not remember specific chromas. From a neurophysiological perspective, the rarity of absolute pitch presents a puzzle. After all, the cochlea of the ear and the auditory cortex share what we call a 'tonotopic' representation of pitch (see chapter 3), which is chroma-specific, similar to a piano. By contrast, we do not have evidence for brain regions that are devoted to specific musical intervals (relative pitch). Why, then, is absolute pitch so rare?

One view, which has received increasing support, is that through the course of development, relative pitch increasingly 'takes over' in humans. Consider the typical musical environment for a human. We usually hear common songs repeatedly such as 'Happy Birthday' sung by different people and in different keys. In order to learn such culturally significant melodies, the listener must recognize similarity across instances based on relative information when absolute information (the constituent pitches) varies. Neuroscientific data suggest that the brains of absolute pitch possessors process pitch differently than do other brains. With respect to anatomy, musicians with absolute pitch have an asymmetry between the left and right hemispheres that prioritizes the left hemisphere. This asymmetry is found in an auditory association area known as the planum temporale, positioned towards the rear (posterior) of the temporal lobe (Schlaug, Jäncke, Huang, & Steinmetz, 1995). This finding is consistent with other research suggesting that the *lateralization* (that is, the degree to which a neural function is exclusive to the left or right hemisphere) of musical pitch varies with training.

We may infer from lesion/function deficit studies that the processing of musical pitch is predominantly a right-hemisphere activity. Since we know that pure pitch identification is a function of the primary auditory cortex, it is not surprising that damage to this region results in loss of simple pitch identification. However, whereas for most people musical pitch is processed predominantly in the right hemisphere, with the left dominant for language, trained musicians can show left dominance also for music (Bever & Chiarello, 1974). Thus the left hemisphere may be dominant for auditory processing that involves the application of internalized categories, such as absolute pitch. Yet another, more recent proposal is that the right hemisphere is optimized for analyzing the spectrum of a sound (which yields its pitch, see chapter 2), whereas the left hemisphere is optimized for perceiving rapid temporal fluctuations (common in speech; Zatorre, 2003).

Key

One of the most striking characteristics of musical pitch processing is the tendency to categorize pitch with respect to the surrounding tonal context (Krumhansl, 1990). In fact, there seem to be specific regions that serve the recognition of the sequences of notes that make up a scale. A recent study using fMRI attempted to identify brain regions responsible for detecting when a single tone 'pops out' based on its divergence from a tonal context (Janata, Birk, van Horn, Leman, Tillmann, & Bharucha, 2002). The researchers simply inserted tones that did not belong to the key, such as an F♯ in the key of C major. The study found consistent activation in the superior temporal gyrus (see Figure 4.2) – an area of general importance for pitch perception (Peretz & Zatorre, 2005), more so for the right than left hemisphere. However, the main focus of the authors was on what part of the brain 'tracks' changes in tonality, which they emphasized as being the rostromedial prefrontal cortex – an area just behind the center of one's forehead.

Other research concerning the way in which tonality makes certain pitches 'stand out' from the rest has applied the ERP technique. This research has revealed changes in electrical responses when a pitch does not match the overarching scale context. Interestingly, more recent work has shown that the brain responds to out-of-tune notes (which do not require a sense of scale, but only require an understanding of basic pitch categories) more quickly than unexpected notes that are appropriately tuned (Brattico, Tervaniemi, Näätänen, & Peretz, 2006). This finding suggests that the interpretation of pitch within a musical scale may constitute a higher-level cognitive task relative to detecting a mistuned note.

Sequential relationships

It is possible that particular neural processes are reserved for perceiving and/or remembering a pitch sequence as an integrated whole and for registering deviations from this. Much of the case study research reported by Isabelle Peretz and colleagues addresses this issue. In the paradigms used by this group, participants are typically presented with melodic sequences in pairs, in which the second sequence is identical to the first with the exception of one note, which may be altered in pitch or duration. Findings from this research suggest that damage to the right temporal lobe selectively disrupts the ability to detect deviations in melodic contour (Liégois-Chauvel, Peretz, Babaï, Laguitton, & Chauval, 1998; Peretz, 1990). Similarly, imaging techniques suggest a separation in the regions where melodies are recognized by overall pitch contour and where they are recognized by sequences of specific interval relations alone. The former tends to be dependent on a right-hemisphere activity, while the latter seems to involve regions in both hemispheres (Peretz & Zatorre, 2005). Peretz (1990) interpreted this finding as supportive of the view that the right hemisphere is involved in more 'holistic' processing whereas the left hemisphere is specialized for finer details.

Integration

As we opened the chapter with a case study, we conclude our discussion in a similar fashion. The case of Rachael Y., described by neurologist Oliver Sacks (2007), serves as a poignant example of the importance of our ability to integrate rich musical works into Gestalt wholes, an ability most of us take for granted. Rachael Y. was an accomplished middle-aged composer and performer when she suffered severe injuries to the head and spine following a serious car accident. (The locations of the injuries were not specified in Sacks's account.) Upon recovering from a coma that lasted several weeks, she found most abilities such as speech to be intact, but noticed a change in her perception of music. She described her experience of the first piece she heard following her recovery from the coma (Beethoven's opus 131):

> When the music arrived, I listened to the first solo phrase of the first violin again and again, not really being able to connect its two parts. When I listened to the rest of the movement, I heard four separate voices, four thin, sharp laser beams, beaming to four different directions. Today, almost eight years after the accident, I still hear the four laser beams equally . . . and when I listen to an orchestra I hear twenty intense laser voices. It is extremely difficult to integrate all these different voices into some entity that makes sense. (p. 113)

As we have discussed, melody is not perceived in isolation. It does not only require the perception of a horizontal sequence, but the complex interweaving of all the parts including the vertical axis (harmony). Having lost the basic ability to integrate the many rich parts of music into a coherent whole, Rachael Y. experienced music as quite unpleasant and chaotic, requiring 'a great cognitive effort to hold the strands together' (p. 116).

Coda

In many ways, the perception of melody is central to our experience of music. GL's case illustrates the dramatic loss of ability to recognize the storehouse of tunes we accumulate throughout our lives. Without the ability to conceptualize tone sequences as gestalts, and to hear the tones as part of a coherent tonal system, music becomes meaningless. At the same time, you may have noticed that the term 'melody' in this discussion has almost exclusively referred to pitch. It is common for those in music cognition to use the term melody in a way that excludes rhythm. The assumption here is that rhythmic relationships contribute independently to musical experience. But do they? We consider this, and other issues related to rhythm, in chapter 6.

Notes

1 See also Satoh, Takeda, and Kuzuhara (2007).
2 This anecdote exists in various versions and is attributed to many different composers, including Mozart. The authors were unable to find a reliable print source and assume the often-told tale to be apocryphal!
3 DeWitt and Samuel's (1990) extensive paper reports a series of five experiments (four using melody lines and one employing chords). The findings of the five studies taken as a whole are more complex than reported here, and the interested reader is referred to the original paper for more detail.

4 Not all participants heard exactly the same percept. Several examples of how the pattern is perceived are shown in the liner notes to Deutsch's (1995) CD. Some participants hear only a single pitch (same pitch 'heard' in both ears) so that two tones are not even perceived. Despite some variation in responses, none of the reports correspond to the actual notes played in each ear.

5 As might be expected, responses to probe tones are highly related to the key suggested by the preceding context, and responses are more distinct for keys that are more distant according to the circle of fifths (see Figure 5.4). For instance, the response to an F# probe tone is much higher (indicating greater stability) after hearing a context in the key of F# than the key of C. Along with Mark Schmuckler, Krumhansl devised a mathematical algorithm that can be used to determine a listener's perception of key by using responses to a series of probe tones (described in Krumhansl, 1990, Chapter 4). This algorithm has been highly reliable and is used often in the literature (see e.g., Temperley, 2001).

6 Perception of musical time

A 1997 study by Mangione and Nieman, which appeared in the *Journal of the American Medical Association*, examined over 500 physicians in training and medical students on their ability to identify common irregularities in recordings of human heartbeats. An unexpected secondary finding was that doctors who played a musical instrument were more accurate at identifying cardiac events than doctors who had no musical training. Cardiac auscultation is one of the most difficult diagnostic skills that doctors must master. 'In 0.8 seconds, you have four or five acoustic events at the threshold of audibility. You need to be able to separate them, and pick them up as a pattern,' explained Dr. Mangione when interviewed about the study (*New York Times*, August 2, 1997). While a direct causal relationship cannot be drawn, it is possible that musical training hones one's ability to hear a rhythmic pattern against an imagined pulse and to detect deviations from it. In this chapter we explore this mysterious 'sense of beat,' a fundamental but complex musical ability that eludes simple explanation.

As we will see, musical timing involves a sense of beat and much more. Time gives life and structure to music. Consider what music would be like if every pitch of a symphony or song was sounded at once! This rather extreme example highlights the importance of time, if only as a way of separating notes from each other. But the succession of notes in time clearly also matters – a scrambled version of a tune as simple as 'Mary Had a Little Lamb' would be unrecognizable. Moreover, we know that tone durations also matter. Consider the first several notes of 'Mary Had a Little Lamb' and 'The First Noel.'[1] When sung in the same key, each melody has the same succession of pitches; it is only the timing of the pitches that differentiates them. Moreover, in music with little or no variation in pitch, emotion can be conveyed in the timing of sounds produced. For instance in drumming, performers can express a range of emotions simply by varying produced temporal patterns (Laukka & Gabrielsson, 2000). In fact, one might be tempted to speculate that the temporal domain is even more important to music's evolutionary history than pitch. Some have suggested that music-making may have served an adaptive purpose for our ancestors based on its tendency to encourage synchronization, whether on the level of groups (Huron, 2003), or in mother–infant relationships (Dissanayake, 2000) as we will discuss in chapter 9. In any case, it is important not to discount the role of rhythm and timing in music.

This chapter is structured around the idea that musical time is multi-faceted. Moreover, we propose that it is this diversity that makes musical rhythms so compelling. First, we discuss research on the formation of *rhythmic patterns* and the way in which these patterns help us to better comprehend and enjoy music. Second, we consider another aspect of musical time: *tempo*, or the rate at which a piece of music is performed. Though conceptually independent of rhythm, there is a complex interplay between rhythm and tempo. Finally, we consider the role of another form of musical time – *meter* – that is associated with perceived regularity in musical time. Following these sections we review possible sources of rhythm, including rhythm's neural bases.

Rhythmic patterns

What, exactly, is rhythm? One of us (PQP) attended a workshop on rhythm held by the neuroscientist Aniruddh Patel. As part of the workshop, the participants (all researchers in music cognition) offered their definitions of rhythm. Nearly every definition was strikingly different! The diversity of our responses highlights an important point. In comparison to a construct like tonality, for which we have a reasonable degree of consensus, the concept of rhythm resists simple definitions.

The simplest way to define rhythm, which we use here, is that it is the time pattern created by notes as music unfolds over time. More specifically, rhythm is a set of time-spans that elapse between note onsets. Ultimately there is more to rhythm than this but we will address this later. Importantly, it is *onsets*, and not note durations, that determine rhythms. Take, for instance, the melody shown in Figure 6.1. Though the version in 6.1a has many durations that differ from the melody shown in 6.1b, both rhythms would sound as if the same. Duration is important in music because it can add emphasis to certain notes (e.g., the note beginning 6.1a); however, durations are not the primary generator of rhythms.

An important fundamental characteristic of rhythms is that they are based on relative time rather than absolute time. Absolute time is time as indicated

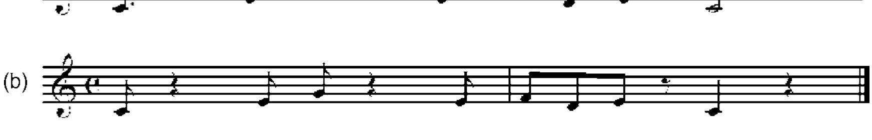


Figure 6.1 Two melodies comprising tones of differing durations that yield the same rhythm based on the timing of onsets.

by a stopwatch, a time-span with no comparison. However, in music, rhythms are supposed to remain constant even when *tempo* (the rate at which music unfolds) speeds up or slows down. Rhythm thus cannot depend on absolute time because the absolute time of every note changes when tempo changes. The fact that rhythms are based on relative time leads to the conceptualization of rhythmic relationships as ratios. The ratio formed by two adjacent time-spans is their *serial ratio* (Jones, 1976). Note that there is a similarity to melody (see the previous chapter). In both cases, the experience of music is thought to be based on perceiving relationships rather than properties of sound heard in isolation.

We now turn to a paradox. Although rhythms are serial ratios, not all serial ratios sound rhythmic! We all know of rhythms that can be difficult or easy to comprehend or reproduce. Others are more complex. We also know of people who attempt to produce rhythms that end up sounding 'un-rhythmic.' Fortunately, serial ratios can help predict how complex (or even un-rhythmic) a musical pattern will be. Simple rhythms come from simple ratios, which are those that can be reduced to an integer value. For instance, if one time interval is exactly half the length of an adjacent interval, regardless of their order, that ratio can be expressed as a 2:1 ratio. This is the so-called 'swing ratio' (though in practice it can be quite variable). However, if one time interval is 500 milliseconds and the adjacent interval is 157, the ratio of 3.18471 . . .: 1 is considerably more complex. The simplest ratio is a 1:1 ratio, followed closely by 1:2, 1:3, et cetera.

The role of rhythmic regularity has been thoroughly explored by Mari Riess Jones, one of the most influential figures in the study of musical rhythms, and her colleagues. A series of experiments explored the way in which rhythmic regularity allows the listener to detect the timing and/or pitch of a forthcoming note. The logic of these studies is that the listener uses rhythms to target attention to forthcoming points in time (Jones, 1976). Rhythmic regularity, in fact, may even help listeners hear pitch (Jones, Moynihan, MacKenzie, & Puente, 2002). Jones et al. used a memory paradigm introduced by Deutsch (1972). The listener is presented with a 'standard' tone to hold in working memory while a series of subsequent 'distracter' tones are presented. Following the distracter tones, the listener hears a final tone and determines whether the pitch of the final tone matches or does not match the initial tone. Listeners were better able to judge matches between the first and last tones based on pitch when the intervening tones were temporally regular than when the intervening tones were irregular.

Here we run into another paradox. Rhythms are simpler when adjacent intervals form interval ratios. However, if one were to analyze the actual performed ratios in recorded music, one would find few such ratios, yet the music can sound quite 'rhythmic.' Chapter 11 (on music performance) will discuss why it is that performers might produce rhythms in this way; here we consider why the listener may hear simplicity in the face of apparent complexity. A possible answer lies in the phenomenon of *categorical* *perception*, a tendency to treat a range of values along a physical continuum as if they were the same until one reaches a point at which the percept abruptly changes. For example, one might present listeners with a series of speech sounds that gradually change from /p/ to /b/. Even though the physical change is smooth and incremental, listeners commonly report hearing a sudden shift – or 'boundary' – at which the repeated presentations of /p/ 'turn into' /b/. Categorical perception of speech is well documented from an early age (see Jusczyk, 1997, for a review). Does categorical perception help us hear rhythm regularity even when presented with variably timed performances?

Clarke (1987) performed a series of simple experiments to test whether listeners have categorical perception for rhythm. Participants were asked to listen to 10 short musical items of five or six notes, regularly timed. Following this initial context, three test notes were played that formed a range of serial ratios between 1:1 and 1:2. Clarke found that people tended to hear either 1:1 or 1:2 even if the timing of the notes actually formed a more complex ratio. Listeners perceived a categorical boundary separating these ratios. Similarly, listeners had difficulty discriminating intervals that did not cross the perceptual 'boundary' between 1:1 and 1:2. Clarke proposed that the more complex ratios were interpreted simply as the result of 'expressive information, or perhaps accidental inaccuracy' (p. 30).

One of the complexities of rhythms is that they can exist at multiple levels. The definition of rhythm described earlier (times between note onsets) can be thought of as rhythms existing on a 'small' time scale. But rhythms can also exist across longer time-spans, and listeners can pay attention to smaller or larger spans, and within limits listeners can choose to attend to one or the other (Jones & Boltz, 1989). These larger time-spans can be created by patterns of emphasis in melody, such as the use of accents (Jones, 1987). An important implication of these larger time-spans is that they create perceived boundaries in music that help the listener to organize cognitively what he or she hears into musical phrases (also called 'groups', cf. Lerdahl & Jackendoff, 1983).

The influence of musical phrase structure on the perception of timing was addressed in a study by Sloboda and Gregory (1980) that borrowed a paradigm from psycholinguistics, called 'click migration.' The experiments on click migration began with an attempt to test the hypothesis that the hypothetical structural properties that were presumed to underlie surface forms in language were psychologically real, rather than convenient abstract representations. Take this example: 'That he was happy / was evident from the way he smiled.' The transition point between first two phrases is marked by a '/'. Participants were asked to indicate with which word a superimposed 'click' was simultaneous. If the click was produced with the fourth word 'happy,' it tended to be heard later than its physical time of occurrence. If it coincided with the word 'was' it tended to be heard earlier. Thus, the click tended to 'migrate' to a natural structural boundary in the speech sequence. The phrasing, a semantic/cognitive aspect of the word sequence, influenced the auditory experience of hearing the click.

In music, Sloboda and Gregory found the same phenomenon. First they presented their participants with musical phrases, the tones of which had been generated by a computer to be of exactly equal length. The tone sequences were structurally identical, yet a click placed on the fifth note (in sequences in which the phrase ended on the fourth note) tended to migrate backwards toward the previous phrase ending. Further, the phenomenon was more striking when the phrase boundary was marked by a longer tone than that with which the click was actually associated. Clearly, musical hearing is more than the mere perception of the physical stimulus.

Tempo

Conceptually, rhythm and tempo are independent of each other. Whereas rhythm refers to relative time, tempo refers to absolute time. More specifically, tempo concerns the speed at which rhythmic patterns unfold.

Tempo is typically considered to be the rate of the 'beat' – a time-span associated with the rate at which a listener will tap his or her foot. Unfortunately this definition is rather too simplistic. In general, people typically try to maintain the percept of a beat that is in the vicinity of 600 milliseconds (which is about 100 beats per minute) even when tempo fluctuates (Parncutt, 1994). Consider a case in which a melody is repeated and is continually sped up after each repetition. Let's say that during the first repetitions, quarter notes (crotchets) occur every 600 ms and so they establish the perceived beat. As the music speeds up, the perceived beat speeds up with each repetition for a while. But then we get to a point where the durations of half notes (minims) near 600 ms. Now the listener may start hearing the beat as being 600 ms again, only now based on the timing of half notes rather than quarter notes. Thus, in a sense, the beat can only speed up so much. Note, however, that when such a perceptual readjustment occurs, the listener is not fooled into thinking that the music has slowed down. In fact, the music can sound quite fast even when the beat may 'slow down' as in this example. Similarly, when a Dixieland Jazz group enters 'double time' (literally doubling the speed), a listener is likely to keep tapping her foot at the same rate, but will easily hear the change in tempo. Thus, though beat and tempo are related, their relationship is complex and not always equivalent.

Why do people cling to a beat that is around 600 ms? One possibility is that the beat, more so than perceived tempo, is linked to the rate at which we like to move in general. Specifically, one's preference for the rate of the beat is highly linked with the concept of spontaneous tempo, first discovered in the landmark research of Paul Fraisse (Fraisse, 1982). Fraisse found that individuals are surprisingly consistent when asked to simply tap at a rate that is comfortable. This behavior varies across individuals but leads to an overall tendency for people to prefer tempos with a beat period around 600 ms. Interestingly, this rate has nothing to do with the heartbeat (sometimes presumed to be the source of rhythm), but is quite similar to the rate at which people walk. As one might expect, differences in preferred walking rate do predict differences in spontaneous tempo during tapping! Fraisse also found that when people were asked to tap prototypically 'long' and 'short' durations, they tended to produce durations that were related by a 2:1 ratio – hence a possible motoric source for rhythmic regularity, discussed earlier.

Another issue that is more complex than it seems at first is the supposed independence of tempo and rhythm. Though this independence holds across a range of tempi, it does not hold in all cases. Various research findings have led some to speculate that the perception of rhythms only holds for time-spans ranging from 200 ms (300 beats per minute) to 1000 ms (60 beats per minute) (e.g. Drake & Botte, 1993), whereas shorter durations are heard as clusters of tones (e.g., grace notes), or as subdivisions of rhythms (London, 2004). By contrast, intervals longer than 1000 ms are heard as separate events, and cannot be heard as reflecting an underlying rhythm unless the interval is subdivided by other tones (e.g., when a solo instrument holds a suspended tone that is subdivided by accompanying instruments). Figure 6.2 illustrates the relationship between tempo and rhythm, generalizing across several studies.

Meter

So far we have equated rhythmic simplicity with *isochrony*, meaning equal inter-onset intervals. But many very simple rhythms are nonisochronous. Take the popular rhythm corresponding to the lyrics, 'shave and a hair cut . . .'. People find this rhythm easy to reproduce but the inter-onset intervals differ from each other (e.g., the note associated with 'shave' is longer than the note associated with 'and'). How do people hear regularity (simplicity) in a rhythm like this? One answer is that rhythms are simple when they clearly match a particular *meter*.

The problem with defining meter is that the definition from musical practice tells only part of the story. Those with musical training are probably

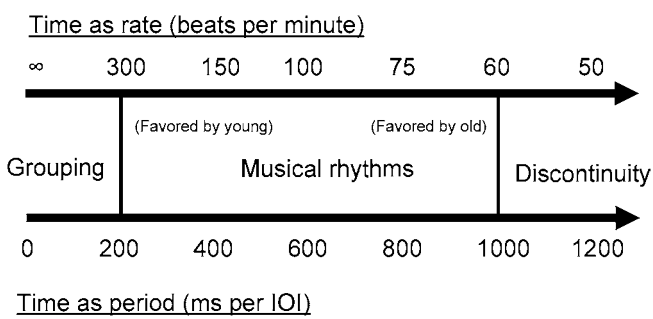


Figure 6.2 The continuum of musical tempi, delineating boundaries in which note events rhythms are typically heard as rhythmic. Copyright © Peter Pfordresher. (IOI = inter-onset interval.)

familiar with the following treatment: the number of beats in a measure or bar. In this treatment, two components emerge. First is the beat, which as mentioned before is an isochronous time-span that is perceptually salient within the musical structure. People often dance to the beat. The second component is the cyclical nature of meter. Meter outlines a recurring period that frames the structure of music. Western music favors cycles that are multiples of twos (marches) or threes (waltzes), Indonesian Gamelan favors fours, while Indian music makes use of very complex interweaving meters.

However, there is another important component to meter that plays an important role in perception. Meter is also conceptualized as a pattern of alternating strong and weak time points. An example of the match between a rhythm and its meter is shown in Figure 6.3. In this representation, called a 'metrical grid,' Xs mark time points on which a note might occur, and the number of Xs determines how prominent or *accented* that time point is. Note that the grid is entirely cyclical, in that Xs at each level repeat regularly. The grid here matches the counting of 4/4 time, but also establishes which sounded notes are more prominent (e.g., the first note and to a lesser extent the third note of the notated melody) and which are less prominent (e.g., the second note and fourth note of the melody). In some cases, a metrical accent can in fact coincide with a silent point; as a result, musical silences can seem conceptually 'loud' (Margulis, 2007). A classic example that she cites occurs in Beethoven's *Eroica* (third) symphony; at one point (measure 280) in the first movement a sudden silence appears at the beginning of a measure – a strongly 'accented' silence.

It is through meter that we get a better understanding of 'the beat' in music. We again draw on the regularity of the grid notation. Notice that there are four vertical 'levels' in the grid, with each level associated with a duration value. Each level should be considered as a possible 'beat' duration. Most people would hear the beat at the quarter note (crotchet) level, hence the time signature. Typically it is thought that the beat lies somewhere in the middle

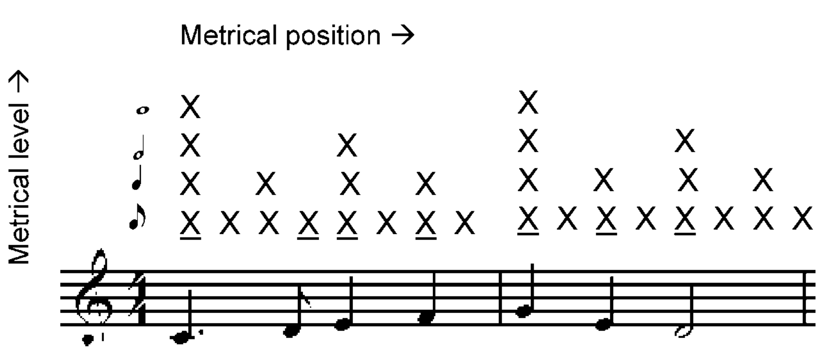


Figure 6.3 A metrical grid. Positions within the grid that are marked by note onsets are indicated by underlining at the lowest metrical level.

(vertically speaking) of the grid, with lower levels constituting subdivisions of the beat.

Why are meter and rhythm thought to be different manifestations of musical time? This is a more sophisticated question than it seems, and not all theorists agree with the dominant view that meter and rhythm are distinct. Arguments for this distinction typically invoke the invariance of meter in the face of varying rhythms. Rhythms rely on the presence of tone onsets. Metrical accents, such as the beat, can exist when no tone is there. Moreover, rhythm and meter can be heard to conflict, as in *syncopation*, an aspect of music that is particularly important to genres like rock and jazz. Syncopation occurs when tone onsets align mainly with weak metrical accents, whereas tones are conspicuously absent at strong accents. Paradoxically, syncopation does not cause the underlying beat to become ambiguous. David Temperley (2000) has pointed out that in rock music the sung lyrics regularly occur just prior to a strong metrical accent and are thus dominantly syncopated in a way that anticipates the beat. 'Let It Be,' by The Beatles, is a good example; the accented syllables (in capital letters) 'MO-ther MA-ry COMES to ME' all anticipate the piano chords that mark the beat. He argues that the use of syncopation in this way draws the listener's attention to the beat, rendering the music more 'danceable.' In general, music that promotes motion (a state often referred to as 'groove' within African-descended genres that dominate current popular music) is often syncopated and often contains events 'on the beat' that actually fall slightly off the beat (Iyer, 2002).

If meter and rhythm are truly separate, where does meter come from? Most would argue that meter originates somehow in the first rhythmic patterns in a piece, which form a context that leads to meter (e.g., Longuet-Higgins & Lee, 1982). However, there is evidence that meter may exist somewhat independently of a particular rhythmic structure. For instance, Palmer and Krumhansl (1990) demonstrated that meter may be generated in the mind based on intentions. They asked listeners to imagine different meters by counting silently, and then presented listeners with simple tones at various times. They found that listeners thought tones sounded 'better' when tones coincided with strongly accented positions, based on the meter that the listener was imagining. This finding was particularly strong for musically trained listeners. Other data suggest that meter persists even when the rhythm has vanished. In one study, Desain and Honing (2003, Exp 2), played a series of 66 rhythms to highly trained musicians and asked them to notate the rhythms as accurately as possible. It was found that the prior presentation of a specific meter (duple or triple subdivision) influenced how musicians conceptualized ambiguous rhythms presented afterwards. Thus, presentation of a specific musical meter may prime listeners to hear metrically ambiguous rhythmic patterns in that meter (see chapter 5 for more on priming).

Several papers have examined the way in which the fit between rhythm and meter can help people remember rhythms and isolate the 'beat.' Perhaps the most influential study in this category is a paper by Povel and Essens (1985). The authors proposed a model that predicts the kind of overarching time-span that might best be suggested by a temporal pattern of inter-onset intervals. Their model focused on the pattern of accents suggested by a temporal pattern. According to their view, a more complex rhythm is one that does not unambiguously suggest a single beat. More generally, this finding suggests that our perception of rhythm is influenced by the degree to which a rhythm can be heard as belonging to a particular meter.

An illustration of the approach advanced by Povel and Essens is illustrated in Figure 6.4. Two rhythms are represented by vertical lines (tone onsets) superimposed above rows of dots that represent candidate time points for tone onsets. Note that both melodies contain similarly complex serial ratios. Thus the perceived complexity of rhythms would only result from how well each rhythm matches a particular beat-based (metrical) structure.[2] As can be seen, the top rhythm clearly matches a structure based on counting in fours (like a 4/4 meter), in that tone onsets always align with group boundaries. However, counting in threes (as in 3/4 meter) does not work as well for this rhythm; thus the rhythm is fairly unambiguous in being binary rather than ternary. By contrast, the lower (complex) rhythm, does not match either organization very well as group boundaries for both organizations can fall on silent points. Povel and Essens found that rhythms like the one shown on the bottom of Figure 6.4 were more difficult for listeners to remember and reproduce than rhythms like the one shown above it.

It should be noted that meter, like rhythm, can range from the simple to the more complex. Certain meters are common and considered to be simple; these meters establish accent patterns that highlight recurring time periods based on roots that are binary or ternary. However, some meters do not reduce to either sort and are referred to as complex. Complex meters often highlight periods of five, seven, or nine. Two of the best known recent examples of complex meters include the jazz piece *Take 5* (a count of five) and the Pink Floyd song 'Money' (a count of seven). In cultures in which these meters are rarely used (like much of Western Europe and North America), such meters can be difficult to conceptualize and might be considered difficult for dancing. In other cultures, such as the Balkans, these meters are quite common. An important set of studies by Erin Hannon and Sandra Trehub (2005a) demonstrated that adults with no exposure to Balkan music had difficulty discriminating deviations from a standard based on a complex Balkan meter, whereas listeners raised in the Balkans had no trouble with this task. At the same time, infants from non-Balkan households were able to perform similarly to the Balkan listeners after only a few weeks of exposure to Balkan meters.

We close the section on meter with a reconsideration of the beat. Earlier we referred to the finding that listeners on average prefer beats that hold close to a 600 ms period. This finding is qualified by age and by musical experience, however. When exposed to the same melody, a younger person would tend to

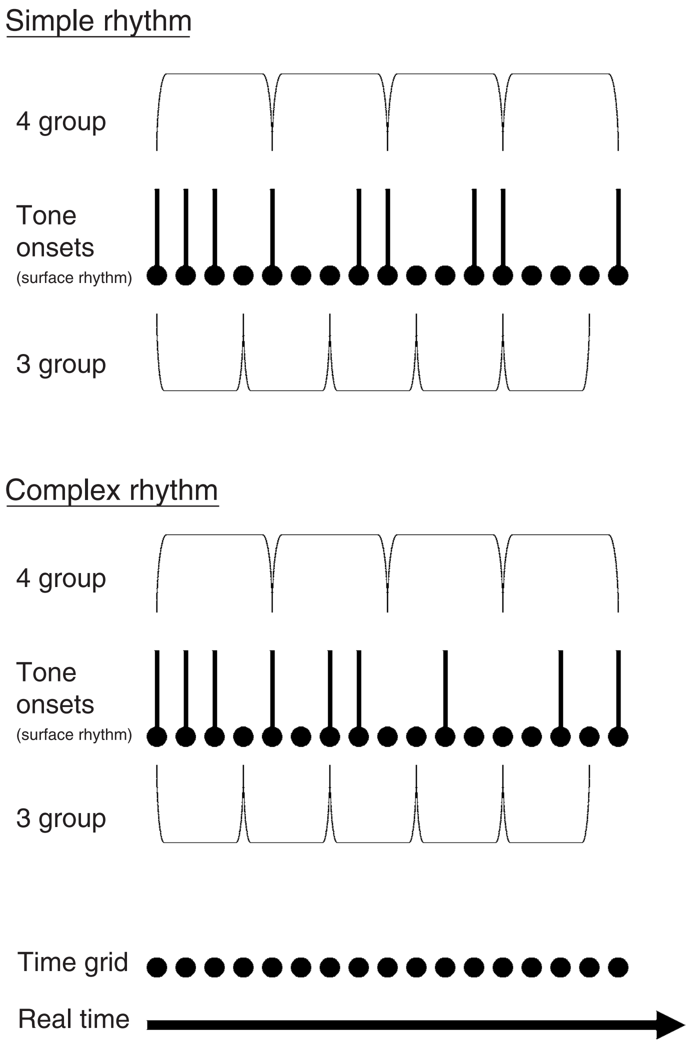


Figure 6.4 Two examples of rhythms used by Povel and Essens (1985). Time in this approach is discrete, represented by dots for candidate time points. The two manifested rhythms (simple, complex) are shown as vertical lines representing tone onsets above dots; dots presented alone represent silence. Brackets above and below represent candidate organizations based on beats that recur every four time points (above) or every three time points (below). Adapted from Povel and Essens, 1985, Figure 5, with permission from the University of California Press.

hear the beat at a faster rater (i.e., a lower metrical level) than would an older person (Drake, Jones & Baruch, 2000) and musicians tend to hear slower beats than nonmusicians (Drake, Penel, & Bigand, 2000). Why is this? For this we turn to possible basic mechanisms that lead to rhythmic behavior.

Sources of rhythmicity

How do we hear rhythms? At present there are two different approaches to this issue. One approach stems from information processing theory that is traditionally associated with cognitive psychology. It focuses on the way in which listening incorporates an internal counting mechanism that essentially computes statistics from musical structure. The other approach has its roots in physics and motor control, and suggests that rhythms are perceived via a process called *entrainment*, which is a process in which one rhythmic pattern achieves and maintains synchrony with another pattern. According to this view, the listener synchronizes internal rhythms with external rhythms, engaging in a kind of internal dance with music. Ultimately both approaches have different strengths and recent theories have attempted to combine the two (e.g., McAuley & Jones, 2003; Pressing, 1999).

The information processing approach has been a dominant force in music cognition, as in psychology overall. Information processing models of timing are typically called 'clock counter' models (e.g., Creelman, 1962). According to these models, tone onsets trigger the initiation of a counting process. During the subsequent inter-onset interval, a series of irregularly timed 'clock ticks' are produced while a second (subconscious) process counts them. There is thought to be variability in the timing of ticks as well as occasional counting errors. Lest this approach sound implausible, consider the fact that 'clock ticks' function like spontaneous neural spikes (action potentials); in fact this may be the neural basis for such a mechanism. Many researchers have been drawn to these kinds of approaches because they match the common intuition that behavior is probabilistic and error-ridden. Moreover, clock-counter approaches have been successful in mimicking the kinds of problems people often have in estimating the passage of time.

Despite these benefits of statistical approaches, many researchers have come to favor a second class of models that arise out of dynamical systems theory in mathematics. These models invoke the concept of *entrainment*. The basic idea is that people are inherently rhythmic, with internal rhythms that adapt to the rhythms of music. We take as a paradigmatic example the model of Large and Jones (1999). The main components of the model are illustrated in Figure 6.5. This model suggests that people time their attention to events in the world by adapting an internal rhythm (a neural oscillation of attentional energy) to note onsets. In Figure 6.5, we see that the peaks in the oscillation match tone onsets. Importantly, the adaptive nature of the internal oscillation can be used to follow music that fluctuates in timing, such as when a performer slows down at the end of a phrase. A further component of the model is the sharpness of the oscillator's peak, which models attentional focus. The peak grows sharper (narrower) when the timing of events is highly regular. Under such circumstances, attention is thought to be narrowly and precisely focused on particular time points. When timing in the world is more variable, attention spreads out in time as does the width of the pulse peak. Note that

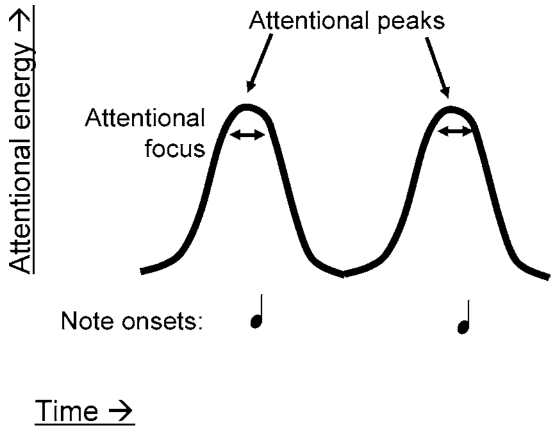


Figure 6.5 An illustration of the entrainment model of Large and Jones (1999). Horizontal double-arrows indicate the width of the pulse (attentional focus). Quarter notes (crotchets) indicate the timing of (acoustic) note onsets. Adapted from Large and Jones (1999, figure 4), with permission of the American Psychological Association.

this 'spread' of attention may explain the phenomenon of categorical perception described earlier.

Entrainment models are attractive because they can predict the way in which listeners adapt to fluctuations of performance timing. Also, unlike statistical models, entrainment models offer a straightforward explanation for why periods that are integer multiples are treated as functionally similar, whereas such changes are considered to be extreme by basic statistical models. The pattern of internal oscillations in such models has led to predictions of listeners' ability to detect temporal fluctuations and to categorize time intervals, as discussed earlier. However, in some ways entrainment models are 'too good' to mimic human behavior, because they do not have any random variability. Such chance fluctuations may be better accounted for by clock-counter models.

The neural bases of rhythm perception

In considering the role of rhythm in the brain one might consider a suggestive fact about the brain: Brain activity is inherently rhythmical. One implication of research on steady-state EEGs is that brain activity often produces regular oscillations, albeit more so during the sleep state than during alert wakefulness. Moreover, some neuroscientists have suggested that the process of learning is guided by theta rhythms in the brain (Givens, 1996). Thus, it seems plausible that our affinity to musical rhythm occurs because our brain synchronizes with rhythms in the world, as in the entrainment approach discussed above.

Recent research suggests that the brain does synchronize with rhythms in the world. This evidence has been yielded by studies using *electroencephalography* (EEG, see chapter 4), due to the high temporal resolution of this technique. The brain responds to patterns of regularly timed events by synchronizing activity with the onsets of these events, and can even respond at the onset of an expected event that does not appear, as in the phenomenon of a ‘loud silence’ described earlier (Snyder & Large, 2005). When a participant imagines tapping regularly, the brain exhibits rhythms in the sensorimotor and parietal cortices that are equivalent in rate to the imagined rhythm (Osman, Albert, Ridderinkof, Band, & van der Molen, 2006).

It also appears as though our brains are predisposed towards certain kinds of metrical organizations. Patterns of oscillating activity suggest a preference for binary (two-beat) over ternary (three-beat) and other kinds of metrical organization (Brochard, Abecasis, Potter, Ragot, & Drake, 2003). Likewise, some neuroscience research supports the presumed separation of rhythm and meter, in that perception of metrical organization is hindered by damage to the temporal lobe in either hemisphere, whereas the temporal lobes may not contribute to rhythm perception (Liégois-Chauvel et al., 1998).

One topic of great concern in research on rhythm perception (and production) is what makes a rhythm simple or complex. We highlight one study that focused on the complexity of serial ratios in rhythm (Sakai et al., 1999). In this study, participants first heard and then reproduced rhythms that were based on integer or noninteger ratios. Functional magnetic resonance imaging (fMRI) was used to measure brain activity as rhythms were held in working memory prior to production. While participants retained simple (integer) rhythms, brain activity was located primarily in motor areas: the motor cortex and cerebellum (cf. Figure 4.2). While participants retained complex (noninteger) rhythms, additional activations were found in the prefrontal cortex. This suggests that the complex rhythms increased memory load given that the prefrontal cortex (just behind the forehead) is important for the functioning of working memory and decision making. In addition, complex rhythms were associated with a tendency for activations to change from being predominantly left hemisphere to being predominantly right hemisphere (suggesting a reduction of ‘analytic’ processing for complex rhythms). Thus, the brain may indeed do more ‘work’ while listening to a rhythmically complex piece like Stravinsky’s *Rite of Spring* as opposed to ‘Twinkle Twinkle Little Star.’ It is possible that the difference between conditions may have something to do with people’s familiarity with simple rhythms; other research has found that when people first learn rhythms, prefrontal lobe activation is found but then dissipates with time (Ramnani & Passingham, 2001).

Of course, rhythms in music are part of the overall auditory signal. This brings up two related questions: Does the auditory cortex show any specialization for rhythm, and to what degree do different parts of the brain contribute to rhythm versus pitch perception? With respect to the first question, there is some evidence that the left temporal lobe, more so than the right temporal lobe, is necessary for rhythm perception, although this specialization may be limited to very short time intervals (Samson & Ehrlé, 2003). Patients who have had portions of their left temporal lobe removed in order to prevent seizures have difficulty discriminating short time intervals that cause perturbations of rhythmic sequences. It is important to note, however, that other areas of the brain are likely to contribute to the perception of musical rhythm. In neuroscience research outside the domain of music, two neural centers have been cited as plausible candidates for one's 'internal clock' (see Ivry & Spencer, 2004, for a review). One is the cerebellum, a large bundle of fibers (75 percent of neurons in the brain) at the rear of the brain (cf. Figure 4.2), just behind the brain stem. The other is the basal ganglia, a ring of fibers that wraps around the top of the brain stem. In this context, it is worth noting that Sakai and colleagues (1999) found cerebellar activations in all of their conditions, although the specific locus of activation differed as a function of rhythmic complexity.

The fact that brain regions for rhythm differ from the regions we discussed earlier when we explored pitch has led some to wonder whether melody and rhythm are handled separately by the brain. Pitches and durations are, after all, integrated in the acoustic signal that we hear. Evidence from cognitive psychology is mixed, with some evidence suggesting that melody and rhythm perception proceed independently of each other (e.g., Palmer & Krumhansl, 1987), and other data suggesting integration (e.g., Boltz & Jones, 1986). As one might expect, evidence from neuroscience is similarly mixed.

Despite the fact that identification of a melody involves recognition of both pitch relations and time organization, studies of patients with brain damage suggest that the brain regions involved are different for each of the major components of musical experience. Cases have been reported in which brain damage to the temporal lobes interferes with pitch discrimination but not time perception (e.g., Liégois-Chauvel et al., 1998), as well as damage that hinders rhythm discrimination but not pitch perception (e.g., Peretz, 1990). Thus someone may be able to identify a melody but have no sense of the rhythmic pattern of the tune. Note that this is a more extreme deficit than the ubiquitous friend who can't dance; It would mean not being able to distinguish 'The First Noel' from 'Mary Had a Little Lamb.' Thus, brain injuries can lead to very specific losses that affect some aspects of musical skills while leaving others intact. This also goes to show that musical ability is complex and multidimensional, drawing on a host of skills in various domains.

At the same time, an fMRI study found overlapping brain regions (primarily right auditory cortex) when listeners tried to detect changes to pitch structure or rhythm in melodies (Griffiths, Johnsrude, Dean, & Green, 1999). In fact, the only difference between these conditions was a region of activation in the dorsal cerebellum during rhythm but not pitch detection.

Other research even suggests that the cerebellum may be active during pitch perception (Petacchi, Laird, Fox, & Bower, 2005). Overall, the mixed findings from both behavioral and neuroscientific research suggest that the interplay between melody and rhythm is still not fully understood.

Coda

Rhythms are intrinsic to the way we move and interact in the world, and are inarguably a deep part of musical experience. Rhythms shape time (cf. Epstein, 1995). Though time is commonly thought to be one-dimensional, like an arrow, rhythmic structure alters this state, leading to a temporal continuum rich with cyclical and hierarchical structures. Theorists have argued that different forms of temporal organization run concurrently as music unfolds, creating sources of tension at times that make music more pleasing and interesting when one is listening to it. Recent theories in psychology suggest that musical rhythms engage us because of rhythms inherent in the listener. As such, listening to rhythms brings about a kind of internal 'dance' between the listener and the music, characterized by figural patterns (rhythm), speed (tempo), and an underlying cyclical structure (meter).

Thus far we have focused our discussion primarily on the musical surface, and on melodic and rhythmic aspects of music. In the next chapter, we turn to ideas about the kind of abstract representations of music that the listener may construct of a musical work.

Notes

1 We are indebted to Caroline Palmer for this example.
2 There is a second component in their approach that we have left out for the sake of simplicity.

7 Analysis and cognition of musical structure

Many a curious child has taken apart a mechanical toy to see what parts are inside it, how they interlock, and what makes them work. Listeners often do the same with pieces of music. Careful listening involves determining how a piece of music is constructed, how the parts fit together, and what makes them work as a whole. It is easy to overlook the complexity of this process. Analyzing a thing in motion, with all its dynamic complexities, poses a formidable task. It takes a trained eye to identify patterns in the movements of a gazelle bounding through a forest, or a flock of birds in flight. Similar challenges are true of listening as the listener hears a succession of tones that each only exist for an instant in time, as a composition is unfolding. Furthermore, the systems that regulate perception and memory are limited with respect to the amount and type of information that can be processed (Snyder, 2000).

Although 'analysis' is often equated with musical expertise, it is important to note that any listener, to some degree, can take part in 'analytical' listening. Although not all listeners are musically trained, most at least grasp sufficient basic patterns or relationships to feel an overall sense of coherence in a new piece of music, to form general expectations about what might come next, to feel somewhat puzzled or frustrated when the music strays dramatically from what they expected, to sense when a phrase (or a whole piece) may be nearing the end, and whether the phrase or composition comes to a satisfying close or not. All the while, they are also organizing the incoming tones into some sort of meaningful array and into coherent units, such as melody and accompaniment, and separate phrases. This process may occur implicitly (see chapter 5), without conscious awareness, but implicit or not our understanding of musical structure is integral to our experience of music.

Musicians – and particularly musicologists (music theorists) – develop the listening skills and vocabulary necessary to develop such analyses consciously and explicitly. Sometimes this process involves an examination of the written musical score, note by note, to more fully understand the interlocking parts – identifying the themes and motifs and their variations, the sources of unity, the harmonic underpinnings of a piece of music. This is the process that is typically referred to as 'analysis' of music. Importantly, this explicit understanding of musical structure can lead to theories concerning how listeners come to comprehend musical structure while listening.

This chapter provides a general introduction to two of the most influential analytical approaches that have shaped our understanding of music analysis and cognition. These approaches are particularly interesting for the psychology of music, moreover, because they are predicated on an understanding of the psychological bases of listening, composing, and performing. It is worth noting that an analytical theory of music is not always a theory of how the listener or performer conceptualizes music. Traditionally, music analysis has been used to elucidate the structure of music, which may or may not be part of anyone's experience of that music. More recently an increasing number of musicologists, led by the example of Leonard Meyer (1956), have become interested in the basic perceptual and memory processes that contribute to musical comprehension.

Lerdahl and Jackendoff's (1983) *Generative Theory of Tonal Music* concerns how a listener organizes aspects of the music that are hierarchically structured (such as motifs nested within phrases nested within sections of a composition, or beats and measures set within a metrical structure). Similar to the earlier approach of Heinrich Schenker (1935), Lerdahl and Jackendoff's theory seeks to 'strip away' all the nonessential notes to uncover the very core of a musical piece. Their theory identifies four components of music that are hierarchically structured, and identifies a system of 'rules' that listeners intuitively apply in order to understand the piece. Even though listening itself is a dynamic activity, their theory is concerned only with the 'final state' of the listener's understanding (p. 4). It seeks to capture the observer's lasting impression after the gazelle has disappeared into the woods, after the flock of birds has vanished into the horizon.

Narmour's (1990, 1992) *Implication-Realization model* departs from the hierarchical approach of Lerdahl and Jackendoff and focuses instead on the musical surface. This theory, which finds its roots in Leonard Meyer (1956, discussed more in chapter 13), is concerned with how a listener's expectations are formed, for instance by what two successive tones imply may come next, and how this implication is realized, to explain the analysis and cognition of basic melodic structures. As its name implies, this model seeks to capture both the prospective and retrospective aspects of perceiving, structuring, and comprehending melody. Thus it seeks to capture the listener's strategies in melodic cognition while the piece is in motion, along with listeners' expectations and some retrospective analysis.

Due to the technical nature of the topic, it should be noted that the present chapter is different from all others, as it is the only chapter that assumes previous knowledge of music theory on the part of the reader. Whereas every effort is made to explain musical terms and concepts in all other chapters, it would be an unwieldy task to do so in the present discussion.

The Generative Theory of Tonal Music

A central assumption of mainstream cognitive psychology is that people apply internalized rule systems to guide their interpretation of events. When it comes to the interpretation of organized sequences, like music or language, many have claimed that these rules take the form of implicitly learned *grammars*. Probably the most famous proponent of this idea as applied to language was Noam Chomsky. Chomsky's famous generative model of the cognitive foundations of language competence (Chomsky, 1966) asks us to imagine that the sentence forms we hear or read are associated with abstract rule systems that can be used to generate any possible grammatically correct sentence (including transformations of the sentence you just heard or read).

Chomsky's view was fundamental in the formation of one of the theories we focus on here, the Generative Theory of Tonal Music advanced by Fred Lerdahl and Ray Jackendoff (1983). For example, summing up the basic principles of cognitive psychology of musical composition, Lerdahl (1988, p. 233) suggests that in any process of composition a 'compositional grammar' is involved, which could be represented by systems of rules. These rules may express the knowledge and skills of a composer who creates a sequence of musical events, or of a listener trying to make sense of the same sequence. As Lerdahl (1988) explains, 'the listening grammar then generates the mental representation that comprises the "heard structure" of the piece' (p. 234). In so doing, the listener makes use of whatever body of musical knowledge he or she has so far acquired.

Background: Schenkerian analysis

Before describing the Generative Theory of Tonal Music in more detail it is important to describe a historically antecedent musical theory, that of music theorist Heinrich Schenker. Schenker believed that the deep structures of music could be revealed by a method of 'reduction' (which we describe shortly) to bring out the basic structures which grounded the music, conceived as an elaboration of these primitive forms (Schenker, 1954, p. 29). Though their work is thought to have been independently developed, there is a common spirit in Chomsky's and Schenker's work (see Sloboda, 1985, pp. 11–23). Both are interested in deriving abstract structures that underlie complex sequences.

According to Schenker, every recognizable melody has a deep structure or *ursatz*. The *ursatz* is complex, having a melodic line (the *urlinie*) and a bass line (the *bassbrechung*). Schenker proposed that the *urlinie* is built from the descending sequence of third, second, and tonic (E D C in C major), whereas the *bassbrechung* is built on the tonic, dominant, tonic sequence (C G C in C major). Schenker suggested that these deep structures are natural and universal (see chapter 15 for further discussion of universality). As such, the listener's capability of melody recognition may rely on the generic *ursatz*. The tonic triad (first, third, and fifth tones) pattern can also be in the repertoire of *ursatzen* since it is clearly the basis on which the whole scheme is built. If we take the suggestion that the 'circle of fifths' (figure 5.4) is the underlying structure stemming from physics and physiology and on which the tonic triad itself rests, the simple 'tonic/dominant' pair would seem to be the deepest *musical* pattern of all. Melody then must consist in departures from and returns to the *ursatz* defined by the basic Schenkerian structures. This is the basis of anticipation, the most important psychological state for the listener's experience of music.

The reduction of a piece of music to reveal the Schenkerian deep structure involves several steps by which the deep structure is revealed as we strip away or 'reduce' the intervening notes which constitute elaborations of base structures. By proposing that all tonal music compositions are generated from a simple fundamental structure (to which they can subsequently be 'reduced'), Schenker provided a way of isolating the distinctive features of a composition that set it apart from features that characterize all tonal music. Figure 7.1 shows an example of a pitch reduction of the Bach chorale *O Haupt voll Blut und Wunden* in the spirit of Schenker but taken from Lerdahl and Jackendoff (1983, p. 108). We should note, however, that the primary goal of Schenker's theory was to identify a universal *ursatz*, from which all melodies could be generated, a kind of 'top-down' approach. The use of Schenkerian theory to create 'reductions,' as we have described, is closer to the goal of most theories in music psychology and constitutes a kind of 'bottom-up' approach (Lerdahl, 2009).

An overview of GTTM

Similar to the ideas of Schenker and Chomsky, music composer and theorist Fred Lerdahl and linguist Ray Jackendoff (1983) set out to describe principles through which an underlying 'structural description' of music can be derived in their *Generative Theory of Tonal Music* (hereafter referred to as GTTM). Put simply, their theory aims to identify the specific rules by which a listener generates a 'heard structure' from the 'musical surface.' Lerdahl and Jackendoff (1983–84)[1] define *musical surface* broadly as 'the physical signal of a piece when it is played' (p. 229) and *heard structure* as 'all the structure a listener unconsciously infers when he [or she] listens to and understands a piece, above and beyond the data of the physical signal' (pp. 229–230). In the broadest sense, therefore, GTTM offers a 'musical grammar' in the form of an elaborate set of rules that explain how listeners come to hear sounds such as pitches of different durations and loudness, as organized music. The applicability of the theory is restricted to tonal music, and the musical grammar focuses only on hierarchically organized aspects of music. The theory consists of four components, each devoted to one of four types of hierarchical structure (referred to as *Grouping Structure, Metrical Structure, Time-span Reduction, and Prolongational Reduction*), each subject to a set of rules.

Figure 7.1 Schenkerian-style reduction of the Bach chorale *O Haupt voll Blut und Wunden* (Lerdahl and Jackendoff, 1983).

Source: Lerdahl, Fred, and Ray S. Jackendoff, *A generative theory of tonal music*, figure 5.4 page 108 © 1983 Massachusetts Institute of Technology, by permission of The MIT Press.

GTTM's reductions (and in particular, prolongational reductions) are broadly similar to Schenker's in that an elaborate musical 'surface' is 'reduced' to its core constituents. However, the two theories differ in their aims. Lerdahl and Jackendoff view GTTM as a 'psychological' theory that aims 'to find principles of music cognition' and Schenker's as an 'aesthetic' theory whose purpose is 'to illuminate musical masterpieces' (1983–84, pp. 248–249). The theories also differ in their focus. For instance, GTTM focuses far more on rhythmic elements than does Schenker's approach.

A musical grammar must identify the kinds of structures that are 'preferred' or 'not preferred,' and 'acceptable' or 'unacceptable,' within a given rule system. Thus, the rules associated with each of the components occupy an important role in GTTM. In Lerdahl and Jackendoff's (1983) words, 'each rule of musical grammar is intended to express a generalization about the organization that the listener attributes to the music he [or she] hears' (preface, p. x). In other words, their rule systems are designed to suggest a kind of 'structural description' that the listener forms while listening to music. There are two primary rule categories they propose. *Well-formedness rules* characterize a fairly fixed set of rules that distinguish 'acceptable' from 'unacceptable' structures. By contrast, *preference rules*, which are more flexible, characterize the kinds of descriptions that a listener will gravitate to when more than one description is possible. In the original theory there was some ambiguity concerning which preference rules would 'win' in specific situations. More recently, Temperley (2001) has developed an extension of the preference rule approach that features a quantitative rule system that can be used to determine which combination of preference rules best matches the interpretation of a listener.

A notational system is needed to visually represent the 'heard structure' of the music, and the general features of this symbol system are also shown in two reductions of Bach's *C Major Prelude* in Figures 7.2 and 7.3. (These figures are only provided as an illustration of the basic features of GTTM's notation system, and a detailed analysis of this composition can be found in Lerdahl and Jackendoff's [1983] book on pages 260–264.)

Grouping Structure 'describes the listener's segmentation of the music into units of various sizes' (1983–84, p. 231), each nested within the other (thus hierarchically structured) such as motifs within short phrases, within longer phrases, within sections of music, and finally within the whole composition. In GTTM's notational style, these are represented by a series of slurs beneath the musical notation, with the broadest grouping (in which all smaller groups are nested) shown at the lowest layer as shown in Figure 7.2. According to GTTM's formal grammar, listeners group the musical surface into segments or chunks consistent with two sets of rules ('well-formedness' and 'preference' rules) that are specified for each of the components.

Here, for example, are five 'well-formedness' rules for groupings of notes (1983, pp. 37–39).

(1) Any contiguous sequence of pitch-events, drum beats, [etc.] can constitute a group, and only contiguous sequences can constitute a group.
(2) A piece [of music] constitutes a group.
(3) A group may contain smaller groups.
(4) If a larger group contains a smaller group, all the elements of the smaller group must be elements of the larger group.
(5) If a larger group contains a smaller group, the larger group must be exhaustively partitioned into smaller groups.

Figure 7.2 Time-span reduction of Bach's *C Major Prelude* (Lerdahl and Jackendoff, 1983).

Source: Lerdahl, Fred, and Ray S. Jackendoff, *A generative theory of tonal music*, figure 10.10 page 262 © 1983 Massachusetts Institute of Technology, by permission of The MIT Press.

Though these principles may sound abstract, a simple example shows that they are in fact highly intuitive. Imagine the first line of the tune 'Twinkle Twinkle Little Star':

Twinkle twinkle little star, how I wonder what you are.

This line when sung may be considered to be a single musical phrase, and thus a group. At the same time, we can hear two smaller groups within this phrase, one functioning as a kind of musical 'question' ('Twinkle

twinkle little star') and the other functioning as an 'answer' ('how I wonder what you are'). Note that these two smaller groups are bounded by the larger group. It simply does not make sense conceptually to have a smaller group (a subphrase) cross a boundary that is formed by a larger group. Furthermore, this single phrase is one of several that together make up the entire tune, and the two subphrases discussed above can likewise be carved into yet smaller groups, the smallest of which would comprise individual notes.

Preference rules are greater in number and more complex than well-formedness rules, and so we will not enumerate all grouping preference rules here but instead focus on two examples. Generally speaking, grouping preference rules suggest where a listener will typically, but not always, hear a group boundary. One such rule is the rule of *proximity*. As in Gestalt psychology (see chapter 5) this rule suggests that a relatively long time-span between notes will often be heard as a group boundary. Going back to our example, the lengthened note associated with the first singing of 'star,' in comparison to the shorter notes associated with 'twinkle, twinkle little . . .,' would lead a listener to (most likely) hear a group boundary where the lengthening has occurred.

But proximity clearly is not sufficient. One could extend the rule of proximity to the point that every single note is its own group; after all there is a 'boundary' after each note in a melody (disregarding chords). To guard against such counter-intuitive 'preferences' Lerdahl and Jackendoff (1983) also propose a rule stating that listeners prefer groups that are not too short in length (certainly not just one event).

Metrical Structure refers to 'the hierarchy of beats that [the listener] attributes to the music' (Lerdahl & Jackendoff, 1983, p. 231), for instance organizing the heard structure into strong and weak metric beats. These form a hierarchy as they may be organized into several increasing broader layers or levels (for instance if in compound time: beats at the individual eighth note [quaver] level, dotted quarter note [dotted crotchet] level, and dotted half note [dotted minim] level). This metrical hierarchy is represented by two or more rows of dotted markings (with the smallest units [i.e., beats] at the topmost level), which appear above the grouping hierarchy in GTTM's notation system as seen in Figure 7.2. (See chapter 6 for an extended discussion of metrical structure and what constitutes a 'beat.')

The well-formedness rules for identifying metrical structure include (Lerdahl & Jackendoff, 1983, p. 69):

(1) Every attack point must be associated with a beat at the smallest level of metrical structure.
(2) Every beat at a given level must also be a beat at all smaller levels.
(3) At each metrical level, strong beats are spaced either two or three beats apart.[2]
(4) Each metrical level must consist of equally spaced beats.

Again, these seemingly abstract principles lead to more intuitive results, in that they lead to metrical structures that are both regular (with respect to spacing of beats) and alternating (with respect to the prominence of beats). For the listener, the result is that meters frame music with respect to alternately stronger and weaker points in time, an effect that we discuss in more detail in chapter 6. For instance, the melody 'Twinkle twinkle little star' yields the strong sense of alternating strong and weak beats on different syllables, with the first syllable of each 'twinkle' being associated with a strong beat.

Preference rules for meter address the thorny issue of which events are interpreted as strong versus weak as well as the link between these patterns of accenting on the musical surface and the underlying meter that a listener interprets from that music. The link from musical surface to meter is not always straightforward, as we discussed in chapter 6. As with grouping, GTTM suggests many metrical preference rules from which we select two examples. Perhaps the simplest preference rule is referred to as the *event* rule. It stipulates that strong metrical accents should coincide with the onset of notes. Thus a listener is unlikely to hear a strong beat in the middle of the sustained note for 'star.' Another, complementary rule suggests that stronger beats are associated with the onset of relatively longer notes. Thus, one is unlikely to hear a strong beat in the middle of 'star' and is instead likely to hear a strong beat at the beginning of that note.

Time-span Reduction 'establishes the relative structural importance of pitch-events within the heard rhythmic units of a piece' (Lerdahl & Jackendoff, 1983, p. 231). Formally, this is notated as a tree diagram that 'expresses the way in which pitch-events are heard in the context of hierarchically organized rhythmic units' (p. 237). In other words, these reductions allow us to identify the relatively important pitch events (tones) in each rhythmic group, which determines the underlying 'deep' pitch structure. Finding Schenker's notation 'not explicit enough' (p. 112), GTTM borrows the 'tree' notation of linguistics to express the hierarchical nature of reductions. However, they point out that the similarity with linguistic trees is cosmetic; theirs are 'purely musical trees, having nothing to do with linguistic trees except that both express hierarchical structures with precision' (1983, p. 112). The tree notation appears above the musical notation, and can be quite elaborate for complex compositions. Underneath the musical notation, the reduction appears on a separate series of staves as seen in the time-span reduction of Bach's *C Major Prelude* in Figure 7.2.

The basic procedure for notating a time-span reduction takes several steps. One starts with a hierarchy of rhythmic groups using principles sketched in the previous sections. In each time-span, for each level of the hierarchy, one identifies a single chord or pitch event that functions as 'the head' (i.e. is prominent) according to whether it acts as a structural point of resolution within that time-span. For instance, in the first group of four chords in Figure 7.2, the initial C-major chord functions as the head; this is evidenced by the fact that the 'branch' from this chord extends over the branches from all other chords, thereby encompassing the rest. As one moves up the hierarchy to broader time scales, tones that functioned as 'heads' at lower levels become subservient to other 'heads' at higher time scales. Extending our previous example, the 'head' of the first four chords ultimately is itself subordinated by the final chord of the sequence, as evidenced by the fact that the branch extending from the final chord reaches over the branch from the initial chord. The result is a hierarchical structure of dominance/subservience in tonal relationships as segmented by time-spans, with branches showing elaborations of pitch events. (The well-formedness and preference rules associated with time-span and prolongational reductions are complex, and will not be specified here.)

Prolongational Reduction. The function of prolongational reduction is to express 'the sense of tension and relaxation involved in the ongoing progress of music . . . the incessant breathing in and out of music in response to the juxtaposition of pitch and rhythmic factors' (p. 179). Movement away from the tonal center, dissonance, a climactic point of a phrase, are all examples of musical tension while return to the tonic, consonance following dissonance, arrival at the end of a phrase, are usually experienced as relaxing. Diagrammed in 'tree' notation appearing above the musical notation, prolongational branching is represented by tensing motions depicted by right branches, and relaxing motion by left branches. The prolongation reduction for Bach's *C Major Prelude* is shown in Figure 7.3.

Unlike the time-span reduction, prolongations (as the name implies) are not constrained by group boundaries. Thus, whereas the resulting time-span units in a time-span reduction are fairly evenly spaced, units in a prolongational reduction can be of varying lengths and can extend across group boundaries. Similar to Schenker's 'prolongation,' a note or chord may remain in effect across measures or even a whole section, even when not physically sounded at each moment throughout them. For instance, a tonic triad can be heard as suspending across an entire section, including several contiguous time-spans. This is represented by the 'slur' notation in the reductions shown on the separate staves at the bottom of Figure 7.3. Note that in many respects the hierarchical structure of time-span and prolongational reductions are similar, as can be seen at higher hierarchical levels in Figures 7.2 and 7.3. However, at lower levels of the hierarchy these structures diverge; one sees less influence of metrical and grouping structure – hence less influence of 'time' – in prolongational than in time-span reductions.

More recently, Lerdahl (2001) has developed the theory of *Tonal Pitch Space* as a means of quantifying the intuitions behind prolongational reductions. This new theory generates distance metrics among pitches in a tonal system (cf. Krumhansl, 1990) and thus can be used to predict the degree of 'stability' associated with certain pitches (and chords), as well as the degree to which certain pitches 'attract' other pitches to them, as in the way that the tonic note 'attracts' the leading tone to it.

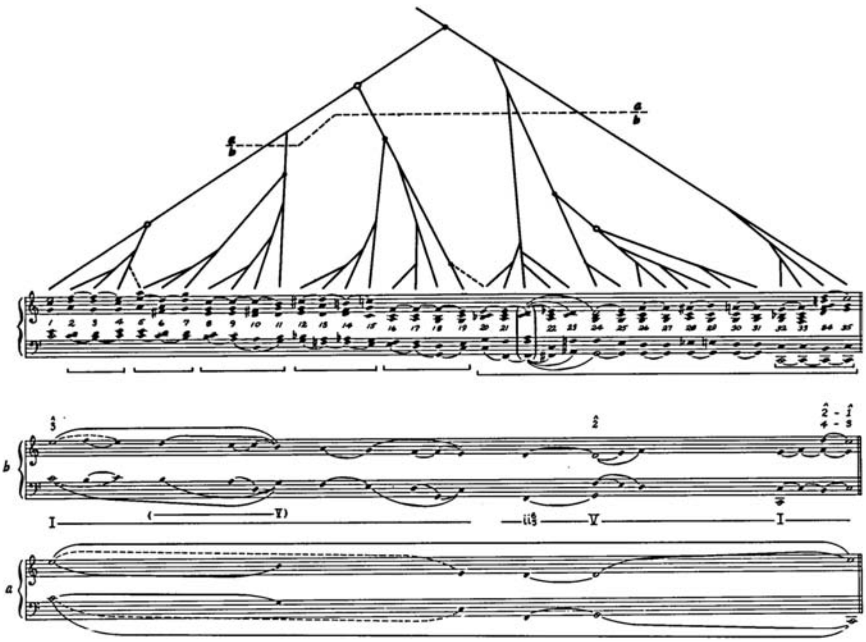


Figure 7.3 Prolongational reduction of Bach's *C Major Prelude* (Lerdahl and Jackendoff, 1983).

Source: Lerdahl, Fred, and Ray S. Jackendoff, *A generative theory of tonal music*, figure 10.11 page 263 © 1983 Massachusetts Institute of Technology, by permission of The MIT Press.

Our brief overview of the theory serves only as an introduction, and has barely captured the complexity of GTTM. In describing the notational system, it is important to remember that it is merely a convenient way of expressing what the theory aims to do – which is to provide a formal description of the musical intuitions of a listener, sorting the 'raw material' at the musical surface of a piece of tonal music into organized 'heard structures.' It is not a theory that focuses on the listener's moment-by-moment response to the tone-by-tone unfolding of a musical work in motion, but rather one that seeks to convey the listener's lasting impressions after a phrase or section or even a whole composition has been presented (i.e., the 'final state' of the listener's understanding [Lerdahl & Jackendoff, 1983, p. 4]).

Empirical support

GTTM has been tremendously influential on the study of music cognition. The website 'Google Scholar' reports well over 1200 citations of Lerdahl and Jackendoff (1983). In some respects, though, its influence has been felt more as an overall perspective than as a specific set of predictions. In comparison to the myriad citations of the theory, relatively few papers have rigorously tested its principles, although the authors have invited empirical tests of their work. This may be due in part to the fact that the original theory was not designed to frame quantitative predictions (Frankland & Cohen, 2004).

Of the empirical work on GTTM, most has been devoted to its claims concerning preference rules. One of the earliest attempts in this regard was Deliège (1987), who demonstrated that perceived segment boundaries are generally well predicted by GTTM rules for music drawn from the standard repertory. More recently, Frankland and Cohen (2004) verified grouping rules with artificially constructed stimuli designed to measure quantitatively the degree to which different principles predict grouping. A major advancement in the quantification of GTTM preference rules is described in Temperley (2001) who instantiated (and extended) these principles in an artificial intelligence system. The theory of Tonal Pitch Space, which evolved out of GTTM's concept of prolongational reduction, has likewise been shown to predict the experience of tension and relaxation in music listening (Lerdahl & Krumhansl, 2007).

The idea that the mental representation of music may be 'reduced' in the sense predicted by GTTM has also seen some support. For instance, influential work by Bharucha (e.g., 1984) has shown that when people hear a tonally unstable note followed by a stable note (see chapter 5), they hear the unstable note as if it were 'anchored' to its more stable neighbor. In other words, the unstable note is treated as somehow incidental. In other research, reductions have been found to predict perceptions of tension and relaxation (e.g., Krumhansl, 1996), closure (e.g., Palmer & Krumhansl, 1987), and memory (Large, Palmer, & Pollack, 1995). At the same time, a problematic result was reported by Cook (1987). He recomposed classical pieces so that they ended in a different key from where they began at the beginning. According to classic Schenkerian views (consistent with the spirit of GTTM), such alterations should profoundly alter the large-scale structure of the piece. Listeners, however, were apparently insensitive to these large-scale alterations. One possibility is that listeners are only sensitive to large-scale structures at a subconscious level that is hard to measure.

Melodic expectations and the Implication-Realization model

We now turn to a different approach to the cognitive representation of music. Whereas the ultimate goal of GTTM is to elucidate the 'deep' hierarchical structure of music, the next model seeks to explain expectations for events at the musical 'surface.' The emotional experiences that listening to melodies brings about are intimately connected with the way in which a sequence of notes fulfills or violates our expectations (see also chapter 14). Although an earlier theory of how these expectations work was suggested by Leonard Meyer (1956), a more fully formalized and widely tested version was proposed by one of Meyer's students, Eugene Narmour (1990, 1992) in his *Implication-Realization model*.

The structure of expectations

The central question Narmour asks is: 'What are the specific, note-to-note principles by which listeners perceive, structure, and comprehend the vast world of melody?' (1990, p. 3). Narmour's Implication-Realization model (hereafter referred to as the I-R model) assumes that a listener's experience of melodic structure is shaped by *expectations of how a melody will continue*. These expectations are influenced by two perceptual systems: one following 'bottom-up' principles that are not dependent on learning and are thought to be universal, and the other applying 'top-down' processes which are influenced by learning and may be specific to certain broad styles or genres of music, or even more specifically to the works of a particular historical period or to a single composer.

Narmour begins with melodic expectations of the broadest sort: the hypotheses that when presented with two musical events of some sort (e.g., two interval patterns, two pitches in succession, etc.), a listener will subconsciously or consciously infer one of two general outcomes:

(a) *Similarity* (or *process*): From the sameness or similarity in successive musical events, a listener comes to form an expectation of more similarity, which can be expressed in the form of the hypothesis (where → stands for 'implies'):

$$A + A \rightarrow A$$

(b) *Differentiation*: Change in successive musical events leads to the expectation of more change, expressed as follows:

$$A + B \rightarrow C$$

What are the dimensions along which sameness or differentiation may occur? Almost any musical dimension, in fact. According to Narmour (1992), these could be 'pitches and durations,' 'melodic intervals or durational patterns,' and musical forms as 'repeated or differentiated units.'

With respect to more specific melodic implications, on its simplest tone-to-tone level, the I-R model accounts for what we expect the third note to be after hearing two notes before it, assuming that the first two notes do not imply a completion. In this context, the first two notes form the *implicative interval*, and the second and third notes form the *realized interval*. (Note that in this context, the word 'interval' should be taken to refer to a *melodic* interval rather than a harmonic interval that is simultaneously played.) Realized intervals that match expectations suggest closure and release, whereas realized intervals that deviate from expectations suggest continuation and tension. According to Narmour, 'it is the dynamic, non-reductive, individuated melodic motion on the note-to-note level that captures music lovers' (1992, p. 331).

Given two melodic tones that are not perceived as completed, the listener generates expectations for the third tone, but not every possible pitch is as strongly implied for the continuation of the melody. The expectancies arising from the implicative interval are, in part, influenced by five 'bottom-up' principles of continuation that have roots in Gestalt ideas about grouping. In our discussion of Gestalt principles of perception in chapter 5, we saw that the Gestalt pioneers were mainly concerned with the visual perception of static figures. However, as Narmour explains, 'in invoking principles of proximity, similarity, and common direction in a temporal art, I hypothesize that rules governing such similarity apply not only retrospectively to perceptual things already seen or heard but also *prospectively* to *expectations* yet to happen' (1990, p. 73, emphases added). Formally, these implicative principles can be set out as follows:

(1) *Intervallic difference*: This principle suggests that after a small interval (five semitones or fewer, a semitone being equivalent to a half-step), the realized interval is expected to be similarly sized; whereas after a large interval (seven or more semitones [or half-steps]), the realized interval is expected to switch to a smaller interval. (Intervals of six semitones, which is an augmented fourth or diminished fifth, are considered neither large nor small, thus lead to neutral expectations.) In other words, intervals are typically expected to be equal to or smaller than a 'perfect fourth' (five semitones), and so when one encounters a large interval ('perfect fifth' or larger) the listener expects the following interval to revert back to the small interval standard.
(2) *Registral direction*: Following a small interval, the listener expects a realized interval that continues in the same pitch direction. That is, an ascending pitch interval that is small generates the expectation of continued ascents to follow. However, after a large interval, the listener expects the next interval to reverse its direction. Thus, a large ascending pitch interval generates the expectation that the next interval will go down.
(3) *Registral return*: A more extreme version of registral direction, registral return suggests that one expects the second interval to bring the pitch back (within two or less semitones [or half-steps]) to where it began at the beginning of the implicative interval.
(4) *Proximity*: Small realized intervals are favored, because these result in pitches that are close (i.e. proximal) to the end point of the implicative interval.
(5) *Closure*: Narmour emphasizes that this principle does *not* refer to the Gestaltist's idea of 'closure' and need not occur at the end of a phrase or section (see footnote 3, Narmour, 1990, p. 151), but refers more broadly to 'syntactic events whereby the termination, blunting, inhibiting, or weakening of melodic implication occurs' (1990, p. 102). Specifically, the implication of the initial interval is inhibited when there is a change in

registral direction, when a large interval is followed by a smaller-sized interval, or both.

Of the principles listed above, the principles of 'registral direction' and 'intervallic difference' play important roles in the I-R model. Moreover, the conjunction of these rules leads to the prediction of a melodic archetype identified by Leonard Meyer as the '*gap-fill principle*' (Rosner & Meyer, 1982). That is, if an implicative interval is smaller than five semitones, a listener expects a continuation of pitch direction. However, if an implicative interval is large the listener expects a reversal of pitch direction, as well as a smaller interval than was heard implicatively. A good example of the latter situation is the opening of 'Somewhere Over the Rainbow,' which begins with an ascending octave that is directly followed by a one-semitone downwards interval.

Note that in certain cases the principles may conflict with each other. For instance, registral return may predict a large interval at the same time as other principles (proximity, intervallic difference) predict a small interval. According to the model, such clashes are not problematic but rather contribute to the tension created by a melodic structure. Further, the I-R model does not restrict melodic implication to aspects of pitch and interval alone. For instance, with respect to closure, factors such as tone duration, meter, and harmony may influence both the location of the implication. These may also determine how weak or strong the expectancy for melodic continuation will be; for instance, Narmour points out that melodic implication can be suppressed by a rest pause, metric stress, or stable tonality.

Bottom-up and top-down expectancies

In Narmour's view, there are two independent 'tracks' operating simultaneously in melodic implication: one which operates on a more automatic and subconscious level (*bottom-up* processes) and one which interprets incoming data more flexibly within the framework of previously learned knowledge (*top-down* processes). So far, we have focused mainly on the 'bottom-up' principles, and in particular the application of elementary Gestalt-like principles to tone-to-tone developments of melodic forms. However, 'top-down' differentiations overlay and modify the bottom-up processes described above. These are the influences of one's knowledge and familiarity regarding musical style as a result of exposure or formal training, and consist of *intra-opus knowledge* (of the particular musical work) and *extra-opus knowledge* (of the style of music). Familiarity or formal training in a specific genre of music or a particular historical period or composer serve as sources from which 'top-down' principles shaping an individual listener's expectations may be derived. Each listener possesses a unique body of style knowledge that influences melodic expectations, and this in part may account for variations among the melodic expectancies of listeners.

Ultimately bottom-up and top-down principles jointly determine musical expectations, the former operating subconsciously and on a local level (e.g., an implicative interval), the latter operating consciously and on a more global level (e.g., pre-existing note patterns). Narmour insists on the independence of the two processes as they shape melodic expectancies in separate and distinct ways:

> In terms of *top-down cognition* . . . a lowly diatonic scale set in quarter notes is a style structure in that, once the listener consciously recognizes its inception, he can easily project its registral, intervallic, metric, and rhythmic continuation and also map out its scale-step conclusion on the basis of his prior knowledge. In terms of *bottom-up perception*, however, the listener also subconsciously, simultaneously, and locally processes each intervallic motion and each registral direction . . . separately since his initial global projection may be mistaken: the major scale he initially envisions from, say, an ascending C-D-E-F might, for instance, actually continue in a quarter-tone fashion after the F.
>
> (1990, p. 9)

We conclude our discussion with two examples of the application of Narmour's model. Let us first illustrate the application of the basic 'bottom-up' principles, or cognitive primitives of the Implication-Realization model, in a simple example shown in Figure 7.4 (from Narmour, 1991, Figure 5, p. 7). The focus here is on the first four notes, or the first three intervals. Intervals 1 and 2 are 'small' according to the I-R theory, and as such one expects a continuation of both registral direction and interval size. However, a surprise comes with the third interval (signified by the exclamation point). Here we have a large interval where a small one is expected. In addition, the large interval continues registral direction and thus we have *differentiation* on one dimension (interval size) and *process* on another (registral direction). This conflict adds to the surprise and – according to Narmour – enhances the listener's aesthetic experience.

Through repeated exposure, the listener can form expectations by way of

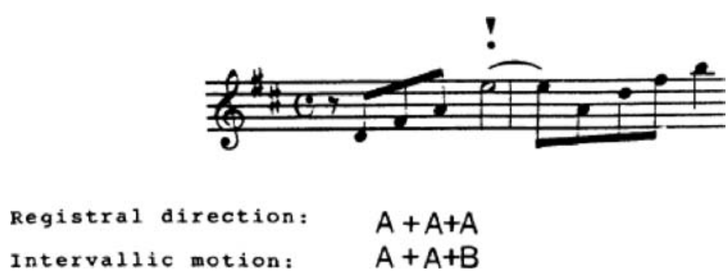


Figure 7.4 'Bottom-up' analysis using the I-R model (Narmour, 1991, figure 5).

Source: Reprinted with permission from the author and the University of California Press.

the top-down system that counteract bottom-up tendencies. This leads to a curious prediction of the I-R theory: Music can surprise our top-down expectations while at the same time fulfilling our bottom-up expectations. An example is shown in Figure 7.5. Here we see a motif that is repeated twice; based on *bottom-up* principles, the listener should expect *process* (a repetition of the motif). However, a different motif follows these repetitions. Narmour argues that, over time, a listener comes to expect the differentiation A + A → B within the *top-down* system in a way that counteracts the A + A → A expectation typical of the *bottom-up* system. The example here does just this. In this example, Narmour points to a clever embedding of themes within themes. Within the 'B' segment identified at the top of the graph there is an embedded theme, B + B → C that mirrors the A + A → B theme at the higher hierarchical level.

The I-R model is complex and consists of many principles to account for expectations, and the above introduction to the I-R model is necessarily simplified. A full explication of the model requires an eight-page glossary of symbols, of which we have not included any. Narmour's various books and articles have elaborated the model we have sketched in several dimensions. Subsequent research has confirmed and extended Narmour's ideas in significant ways, as discussed in the next section.

Empirical support

Like GTTM, empirical tests of the I-R model have primarily focused on the simpler claims of the model, that is, its 'bottom-up' components. These principles appear to account for melodic expectancies quite well, particularly when combined with a component based on the tonal hierarchy that arises from Krumhansl's (1995) research. For instance, when given two tones and asked to rate how well a third tone serves as a continuation of the melody, the responses were in line with Narmour's predictions (e.g., Cuddy & Lunney, 1995). Similarly, when participants were given two notes and asked to

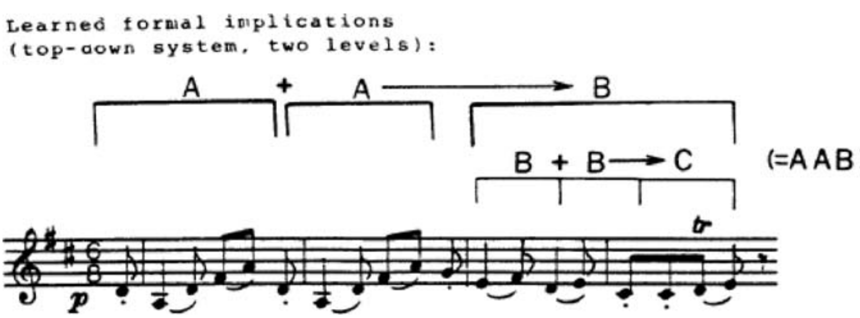


Figure 7.5 The interplay between 'bottom-up' and 'top-down' implications (Narmour, 1991, modified figure 15a).

Source: Reprinted with permission from the author and the University of California Press.

compose the rest of the melody, the third tone they chose in creating their own melodies was well predicted by the I-R model (Thompson, Cuddy, & Plaus, 1997). The level of musical training of participants did not make a difference in these studies, which is consistent with Narmour's assumption that the application of 'bottom-up' principles is not dependent on formal musical training. Further, Krumhansl (1995) and Schellenberg (1996) have shown that these 'bottom-up' expectations can transcend cultural boundaries, and can apply to expectancies for Chinese music and – unlike Krumhansl's tone profile – they can also account for expectances in atonal music (Krumhansl, 1995).

However, in one recent cross-cultural study the I-R model was shown to be limited (Krumhansl et al., 2000). This research investigated expectancies for Sami Yoiks, a style of music originating in Northern Europe and characterized by large pitch leaps. As one might expect, the I-R model, based largely on an assumption that small intervals are favored, did not provide as strong an account of expectancies for Yoiks as it has for other musical styles. Interestingly, the fit of the model was culture-specific. Listeners who were unfamiliar with Yoiks made expectancy implications that were better predicted by the model than did listeners who were familiar with Yoiks. In other words, Yoik-listeners had modified the schemata they used to form expectations for melodies.

Narmour's model has also stimulated further theoretical work, refining and expanding models of melodic expectation. Schellenberg (1996, 1997), for instance, demonstrated that the I-R model could be simplified by removing some predictor variables without loss of its predictive power. His parsimonious 'two-factor model' proposes that melodic expectancy is determined by the principle of pitch proximity and pitch reversal, two ideas articulated by Narmour. Drawing on the work of Narmour (1990, 1992) and Lerdahl (2001), Margulis (2005) proposed a model of melodic expectation, which assigns expectancy ratings ('predictions about degrees of expectedness' [p. 664]) to melodic events and associates them with listeners' experiences of tension across melodies. Aside from providing a rubric for computing the expectedness of melodic events, the model also 'enables the graphical representation of moment-to-moment expectancy-based fluctuations in affect across a melody's course' (p. 665), capturing the dynamic and real-time aspects of listening to music.

Coda

There are many different ways to take apart a mechanical toy or a clock, a poem, or a musical work. The guiding questions we have in mind, the way we inspect the internal structure, the tools we use, and what we choose to focus our attention on, may all lead to different but often complementary discoveries. In this chapter, a few different approaches to understanding musical structure were explored. The Generative Theory of Tonal Music and the Implication-Realization model capture different aspects of the way listeners perceive and comprehend a piece of music, and both have contributed significantly to our understanding of music cognition. However, none of the theorists claim to have set forth a 'complete' or 'finished' theory. Further, the theories do not attempt to 'explain music' but to capture the complex combination of innate and learned responses in listeners' perception and cognition of a musical work. Drawing from an analogy by psychologist Paul Fraisse, Narmour explains: 'gravity does not explain architecture, but architecture is subject to its law; likewise, perceptual laws do not explain music, but music cannot escape their influence' (1990, p. 4).

The theories described here offer a good starting point toward understanding the perception and cognition of a piece of music. However, many questions ensue: How do musical structures give rise to emotions in listeners? And how does music come to have *meaning* for the listener? These questions are addressed later, in chapters 13 and 14. But first, having explored the perception and cognition of melody, rhythm, and musical structure in the previous and present chapters, we turn our attention to the emergence and development of these and other musical abilities in the next chapters.

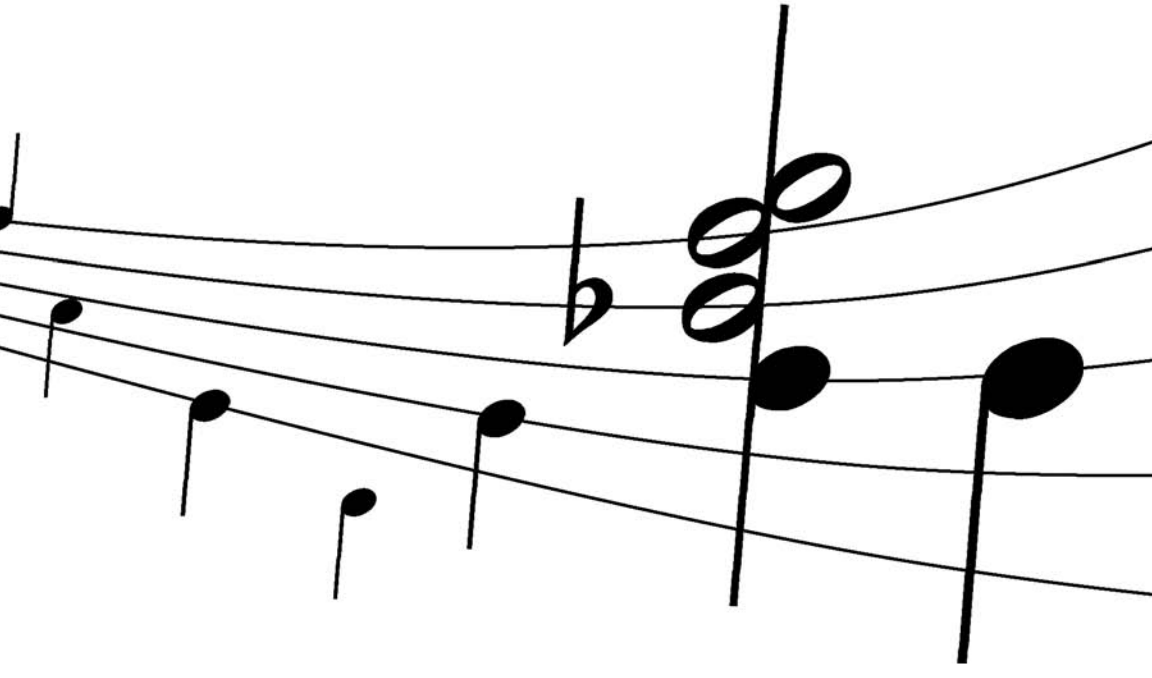
Notes

1 The publication date for Lerdahl and Jackendoff's paper, 'An overview of hierarchical structure in music,' is cited in this chapter as 1983–84. Note that this is a different publication from their 1983 book entitled *A Generative Theory of Tonal Music*.

2 This rule would lead to the privileged choice of the time signatures 4/4 or common time and 3/4, waltz time.

Part III

Development, learning, and performance



8 Emergence of sound and music perception

'Let's start at the very beginning. A very good place to start . . .' So go the opening lyrics of Rodgers and Hammerstein's 'Do Re Mi' in *The Sound of Music* (Wise, 1965). In previous chapters, we have described the auditory system and melody and rhythm perception in the mature adult. In the present chapter, we retrace our steps to the very beginning to explore the emergence of hearing and music perception before birth and in infancy. This chapter is divided into two sections. The first part explores the development of hearing *before* birth and presents some of the main findings in the literature on fetal and newborn responses to auditory stimuli, focusing on responsiveness to the human voice and music. This will require some discussion of what sounds are available to the fetus before birth, and the methods by which we have reached these conclusions. In the second part of the chapter, we explore music perception in the first year of life. Throughout, we describe research procedures in some detail, as readers are often curious as to how studies are carried out during the prenatal period and with young infants.

Part One: Auditory capacities before birth and in the newborn

Perhaps no period of human development is as shrouded in mystery as the months between conception and birth. The course of prenatal auditory development, in particular, is a relatively new area of study and ideas about what the fetus can hear *in utero* have undergone significant changes. In the early 1900s, it was commonly assumed that humans had little or no auditory sensitivity at all before birth. Expectant mothers' reports about movements of the fetus in response to loud environmental sounds were routinely met with skepticism. Gradually, however, scientific papers began to corroborate these anecdotal reports. In an early study, Sontag and Wallace (1935) devised a sort of stylus recording device out of inflatable sacks resting against the abdomen, attached to four pens that recorded changes in the shape of the abdomen. When a wood block resting against the woman's stomach was struck loudly, the researchers noted corresponding movements in the contour of the abdomen. Observing this, they inferred that the fetus responds to external sound or vibration before birth. With new technological advancements, much more precise and reliable methods have been devised to study auditory capacities before birth.

Methods for studying fetal responsiveness to sound

One way to examine the auditory capacities of the fetus is to introduce auditory stimuli into the intrauterine environment and monitor any corresponding fetal responses. Several techniques can be used to present sound to the fetus. In the *airborne* mode, sounds are played via loudspeakers placed near to the mother, usually at a distance of about 1 to 3 feet. In the *air-coupled* mode, sound is transmitted through loudspeakers or headphones placed directly on the mother's abdomen, sometimes cushioned by a foam ring to reduce distortion created by direct contact with the skin. In the *vibro-acoustic* mode, a vibrating device such as a tuning fork or a voice simulator is placed directly on the maternal abdomen.

To determine whether the fetus can detect these sounds, motor responses and physiological changes in the fetus are carefully monitored. Motor responses, such as an increase or decrease in fetal limb movements or a 'startle' reflex, can be observed through real-time ultrasound scanning. Physiological changes, such as a change in fetal heart rate following presentation of sound, can also be tracked using fetal tocographs (heart monitors) and other monitoring devices. Researchers rarely rely on maternal reports of fetal movements as the sole measure of fetal responses. Women may perceive 27 to 75 percent fewer fetal body movements than are observed in ultrasound scanning (Kisilevsky, Killen, Muir, & Low, 1991).

Studies employing some of the methods described above suggest that human auditory functioning begins *about 3 months before birth*. While full-term infants are born at 38 to 42 weeks gestational age (or GA), fetuses already respond to sound by about 25 to 28 weeks, as indicated by an increase in fetal heart rate or fetal movements in response to sound or vibration (e.g., Kisilevsky & Low, 1998). Cortical potentials (rapid fluctuations in brain activity) in response to auditory stimuli have also been detected in preterm infants delivered at only 25 weeks GA (Starr, Amlie, Martin, & Sanders, 1977). The structures of the auditory system, as outlined in chapter 3, are well developed before birth. In particular, the development of the cochlea progresses at a rapid rate: by only 8 to 9 weeks GA the cochlea is fully coiled, and the hair cells begin to differentiate at around 10 to 11 weeks GA (Pujol & Lavigne-Rebillard, 1985). The development of the cochlea is completed by about 30 weeks GA (Pujol, Lavigne-Rebillard, & Uziel, 1990), but cochlear functioning begins some weeks before this, as suggested by the aforementioned studies.

Internal and external sounds in the intrauterine environment

Sounds in the intrauterine environment arise from two sources. *Internal sounds* originate from inside the mother's body, such as the mother's breathing, abdominal gurgles and other digestive sounds (collectively referred to as 'borborygmi'), vascular activity, and heartbeats, as well as similar biological sounds originating in the fetus. *External sounds* originate from the environment outside the mother's body such as external voices, music, ringing telephones, and traffic noise. The maternal singing and speaking voice is both internal and external in source, as it originates internally but is also transmitted as an airborne sound.

Early recordings of intrauterine noise levels obtained with microphones inserted into the uterus yielded internal noise readings of about 72 to 96 decibels (dB) (e.g., Lecanuet & Granier-Deferre, 1993). Seventy decibels is comparable to the hum of a noisy car engine or vacuum cleaner, and 90 dB would be four times as loud (as decibels correspond to a logarithmic scale, as explained in chapter 3). It was therefore assumed that even if the auditory system were already functional, external sounds such as music and voices would be masked or completely obscured by internal sounds. However, early recording methods were not adapted to address fluid impedance. Findings from more recent studies using equipment such as hydrophones (designed to detect sound in underwater environments) suggest a much quieter intrauterine sound environment. These readings range from about 28 to 65 dB, with wide variability varying with the placement of the recording device, time of day, activity level of the fetus, and other factors (Lecanuet & Granier-Deferre, 1993; Querleu, Renard, Boutteville, & Crépin, 1989, p. 412). Twenty decibels is comparable to the sound level of rustling leaves and 65 dB is the approximate loudness of conversational speech – thus the fetus is most likely exposed to a much richer repertoire of external sounds than previously thought.

Quality of external sounds in the intrauterine environment

As might be expected, one important variable affecting the quality of transmitted external sounds is *amplitude*. External sounds must be fairly loud to reach the fetus, as they must pass through the maternal abdominal skin and be transmitted through the uterine tissue and fluid. This can lead to some *sound attenuation* (i.e., *loss* of strength of the auditory stimulus, or the opposite of amplification) so the sound reaching the fetus is weaker than the original source.

In general, external speech presented in airborne mode at a normal conversational level of about 60 dB and external music presented in airborne mode at about 80 dB (a rather loud listening level) can penetrate through the maternal tissue and emerge above the intrauterine noise level (Querleu, Renard, Versyp, Paris-Delrue, & Crépin, 1988; Woodward, 1992). However, the quality of the sound reaching the fetus also depends on the level of internal noise, which varies with the mother's level of activity. Internal noises such as abdominal gurgling following a large meal, or cardiovascular sounds and movement of amniotic fluid during physical activity, may mask external sounds. If the mother attends a concert following a hearty dinner, or listens to music while exercising, high internal noise levels may be competing with music from the external environment!

Another variable mediating degree of sound attenuation is *frequency*. In general, high-frequency sounds are subject to greater attenuation via maternal tissue than low-frequency sounds while lower-pitched external sounds are more efficiently transmitted to the fetus. In effect, the abdominal wall acts as a ‘low-pass filter.’ Therefore, musical tones up to 250 to 300 Hz range (from the lowest keys on a grand piano to just above a middle C) may pass through to the intrauterine environment virtually unattenuated – that is, with little loss in sound energy (Abrams, Gerhardt, & Peters, 1995). On the other hand, high-frequency sounds toward the highest octaves of a grand piano may not clearly emerge from the intrauterine background noise (Querleu et al., 1989).

Several studies have taken recordings from inside the womb in order to assess the clarity of external sounds reaching the intrauterine environment. For instance, Denis Querleu and colleagues (1988) asked French-speaking adults to reproduce the speech sounds they could hear from a recording of voices obtained by a recording device located next to the fetus’ ear. The stimuli used were presented in airborne mode at a conversational level of about 60 dB, and consisted of items from audiometry lists consisting of meaningful and nonsense words (so that participants could not guess inaudible syllables from partially heard words). Querleu and colleagues found that adult participants were able to recognize only around 30 percent of about 3000 phonemes and 22 percent of the consonants. Spoken consonants are higher-pitched than vowels, and thus are subject to greater attenuation than vowel sounds. External speech therefore loses much of its crispness by the time it reaches the fetus.

Similar studies have been conducted using orchestral music. In one study, Woodward and colleagues played Bach’s *Brandenburg Concerto No. 1* presented at 80 dB, and recorded the sound from inside the uterus (Woodward, Guidozzi, Coley, Anthony, & De Jong, cited in Woodward, 1992). Attenuation of the higher frequencies made the music sound more muffled and less brilliant than the original recording. Despite some loss in tonal quality of higher frequencies, however, many of the distinctive qualities of the music – such as rhythm and orchestra texture – were preserved. In another study, for example, an intrauterine recording of a segment of the first movement of Beethoven’s *Fifth Symphony* opus 67 was easily identified by adult listeners (cited in Abrams et al., 1998, Figure 4).

Fetal responses to music

Fetal responses to music have been investigated in only a handful of studies, but there is some evidence that fetuses respond to music in the last trimester (i.e., last 3 months) before birth. For example, changes in heart rate or motor responses have been reliably observed in fetuses exposed to air-coupled presentations of Chopin's *Minute Waltz*, Brahms' *Waltz No. 15*, Gounod's waltz from *Faust*, and Bach's *Organ Prelude* BWV 548 (Olds, 1985; Woodward, Guidozzi, Hofmeyer, Kent, & Warton, cited in Woodward, 1992). Kisilevsky and colleagues also found that the presentation of a piano rendition of Brahms' *Lullaby* evoked heart-rate acceleration at 33 weeks GA and body movements at 35 weeks GA (Kisilevsky, Hains, Jacquet, Granier-Deferre, & Lecanuet, 2004).

To date, it is difficult to say whether there is any correspondence between the specific characteristics of music and the type of fetal responses elicited. Shetler (1989) exposed fetuses to 'stimulative' and 'sedative' music through the air-coupled method. He reported that fetuses responded with faster, more agitated movements to the stimulative music, and softer and more flowing movements to the sedative music. However, it is likely that fetuses simply responded to the broad acoustical differences between these two types of music. Similarly, most researchers have been cautious in drawing any conclusions about fetal preferences for different styles of music. For example, one study showed that heart rate decelerated and frequency of movements was higher in 38-week near-term fetuses listening to a Beethoven piano sonata (*The Tempest*) as opposed to a choral work or rock music (Wilkin, 1995). However, the researcher certainly did not conclude that fetuses have a special preference for Beethoven's music, but that these responses may be due to the 'dramatic acoustic changes,' such as sharp contrasts in volume, tempo, and texture, in the Beethoven piano sonata. Again, it is likely that fetuses are responding to broad acoustical features as opposed to fine characteristics of music.

The possibility that maternal responses might mediate fetal responses should also be considered. Zimmer et al. (1982) played classical and pop music to mothers via headphones placed on the mothers' ears (not abdomens). The researchers found that 34- to 40-week-old fetuses showed an increase in motor activity and a decrease in respiratory movements when mothers listened to 25 minutes of either classical or pop music, compared to the control condition in which no music was played. These effects were stronger for the genre of music that the mother reported that she preferred. The authors conjectured that the mother's liking for a piece of music may affect her physiological state (e.g., degree of relaxation and mood), which in turn may affect the fetus' responses. For instance, if the mother's abdominal muscles are relaxed when listening to pleasant auditory stimuli, this may increase the amount of space in which the fetus can move (Hepper, 1991). The mother's role as a possible mediating influence is important to consider in studies in which the mother can hear the music presented to the fetus. After all, the mother usually hears the music to which the fetus is exposed under 'normal' circumstances outside the laboratory.

Links between prenatal exposure and newborn responses

Maternal heartbeat and voice

A question that is often asked is whether prenatal exposure to particular sounds can have a lasting influence after birth. For instance, are newborns soothed by the sound of the human heartbeat, as it is a familiar sound from the womb? In a classic study, Salk (1961) found that newborns who heard the sound of a heartbeat of about 72 beats per minute (presented for 4 days) cried less and were more likely to gain weight during those 4 days than were newborns in a control group. However, subsequent studies have yielded mixed results. Many do not suggest a special preference for the human heartbeat, as the same short-term calming effects can be elicited by other rhythmic noises with similar acoustic characteristics. For instance, newborn infants respond similarly to the sounds of the resting heartbeat, lullabies sung in a foreign language, and even metronome clicks at 72 beats per minute (Brackbill, Adams, Crowell, & Gray, 1966). It is possible that fetuses are conditioned to respond to heartbeats through prolonged exposure in the womb, but that this response can be generalized to many other slow rhythmic sounds.

Another question often asked is whether newborns are particularly responsive to the sound of their mothers' voices. DeCasper and Fifer (1980) showed that infants under 3 days old already prefer their own mother's voice to the voice of another female. In this study, newborns sucked on a nipple that was connected to sound equipment through a pressure transducer, so that pressure on the nipple could control the audio-track being played. The rhythmic pace of the newborn's sucking produced either a recording of the mother's voice or of another woman's voice. Newborns quickly learned the pattern of sucking that elicited their own mother's voice and produced her voice more often than that of the other female, even though both women were reading the same prose passage. Originally, DeCasper and Fifer concluded that the newborns had learned to recognize their mothers' voices after short-term *postnatal* exposure and only conjectured about the effects of *prenatal* exposure. However, even newborns under 2 hours old already demonstrate a preference for their own mother's voice (Querleu et al., 1984) while 2-day-old newborns did not prefer their father's voice to the voice of another male following 4 to 10 hours of postnatal contact (DeCasper & Prescott, 1984).

Newborns may also respond to familiar passages of prose to which they were repeatedly exposed during the last weeks of their mothers' pregnancy. In DeCasper and Spence's (1986) widely-cited study, for instance, expectant mothers were asked to read one of the following stories out loud every day during the last six weeks of pregnancy: Dr. Seuss' *The Cat in the Hat* (1957), *The King, The Mice, and The Cheese* (Gurney & Gurney, 1965), or *The Dog in the Fog* (a story the researchers created by replacing nouns in *The Cat in the Hat*). Using the same procedure as in DeCasper and Fifer's study, the researchers found that 2-day-old newborns preferred the story that had been read to them while *in utero*, to stories that they had not been exposed to prenatally. Because the familiar story was preferred even if the recording was made by a female that was not the mother, the researchers concluded that newborns can recognize some acoustic characteristics of speech, such as syllabic stress and voice-onset-time of consonants. Indeed, a subsequent study by Kreuger and colleagues suggests that by 34 weeks GA, even *fetuses* may already respond to familiar verse, as indicated by a cardiac response to a rhyme to which they were previously repeatedly exposed (Krueger, Holditch-Davis, Quint, & DeCasper, 2004).

Vocal and instrumental music

Most studies that have investigated transnatal continuity of auditory experience have focused on responses to speech, and relatively few have focused on possible effects of prenatal exposure to music. Using similar methods to the ones devised by DeCasper and colleagues, both Satt (1984) and Panneton (1985) found that newborns preferred a lullaby that mothers had repeatedly sung during the last trimester of pregnancy to an unfamiliar lullaby that the infants had never heard before.

Similar results have been found for voice with accompaniment or instrumental music. Hepper (1991) played the catchy musical theme to *Neighbours* (an Australian television soap opera) to fifteen 2- to 4-day-old newborn infants. When the *Neighbours* theme was played for the first time since birth to newborns whose mothers reported watching the soap opera daily while they were pregnant, the newborns exhibited a decrease in heart rate, a decrease in movements, and tended to adopt an alert state while exposed to the music. Newborn infants whose mothers had not watched the soap opera before did not exhibit any significant changes in heart rate or movements, and very few became quiet and alert upon hearing the music.

The effects of prenatal exposure to music have also been observed both before *and* after delivery. In Wilkin's (1995) study, for example, fetuses were exposed to an audiotape of classical and rock music, played daily from 32 weeks gestational age to 6 weeks after birth. Compared to a group of matched controls that were not prenatally exposed to the audiotape, more fetuses in the experimental group showed heart-rate deceleration to presentation of the audiotape at 38 weeks GA. Newborns in this group were also more likely to adopt an alert state when the audiotape was played at 6 weeks after birth.

Beneficial effects of music for the newborn?

It is interesting to learn that infants respond differently to some sounds to which they were regularly exposed during the last few months before birth. However, it must be noted that this apparent 'recognition' to familiar materials is short term, lasting only a few days to several months after birth (e.g., Fifer & Moon, 1989). As yet, there is no hard evidence that prenatal exposure to music or other auditory stimuli 'teaches' the fetus something that will carry through to childhood and beyond, nor primes them for exceptional ability. These early 'memories' are probably lost in the extensive rewiring of neurons that takes place in the first several years after birth. The main benefits of music listening may be for relaxing the expectant mother, thereby bringing indirect and nonspecific developmental benefits to the fetus.

Similarly, the idea there are lasting benefits of playing classical music to newborns and infants – and in particular, that it can boost cognitive functioning – has not yet been clearly demonstrated in the scientific research. This popular notion, and the way it came to be associated with *the Mozart effect* (or the alleged enhancement of cognitive skills by listening to classical music, particularly the music of Mozart), has an interesting history. The origin of the so-called 'Mozart effect' can be traced back to a brief paper by Frances Rauscher, Gordon Shaw, and Katherine Ky which appeared in *Nature* in 1993 (followed by a longer 1995 article by the same authors). The study showed that listening to Mozart's *Sonata for Two Pianos in D major* (K. 488) for 10 minutes enhanced adult listeners' performance on standard spatial tasks, compared to listening to a relaxation tape or to nothing at all. Rauscher and colleagues found that those who had listened to the Mozart sonata immediately before completing three spatiotemporal reasoning tests taken from the Stanford-Binet intelligence scale performed significantly better than those in the other conditions. However, the effect was short term, 'washing out' after about 10 to 15 minutes.

Rauscher and her colleagues were cautious in interpreting the findings of their study and took pains to point out its limitations. However, when findings of the study were disseminated in the media, where they were dubbed the 'Mozart effect' (a phrase which the original researchers never used), the reports often seemed to imply or even directly claim that listening to Mozart can boost intelligence or IQ. The fact that the original study focused on a specific type of intelligence (spatiotemporal reasoning), that the effect lasts only about 10 to 15 minutes, and that subsequent studies have usually failed to replicate the original findings (Chabris, 1999) were often omitted.

Other explanations for the original findings have received far less attention in the media. For instance, the original interpretation of the Mozart effect was that listening to Mozart stimulates neural networks that are also used in spatial reasoning. However, Nantais and Schellenberg (1999) pointed out that the three conditions originally used (Mozart, silence, relaxation tape) did not merely differ with respect to their ability to stimulate this network, but with respect to the boredom they caused in the listeners, which was substantially less for Mozart's music than the other conditions. Rather than stimulating a specific neural network, subsequent studies suggest that Mozart's music simply caused more arousal in listeners (Thompson, Schellenberg, & Husain, 2001) and this effect may therefore be brought about by a range of other stimuli including music by other composers, music of nonclassical genres, and even recorded prose passages. It has long been known that moderate amounts of arousal can facilitate performance on many tasks.

One of the unexpected outcomes of this research is that – despite the fact that infants were not the focus of the original or subsequent scientific studies from which the idea was derived – the Mozart effect became strongly linked in public perception with the idea of wide-reaching benefits of classical music for infants. Bangerter and Heath (2004) analyzed 478 articles related to the Mozart effect appearing in the top 50 American newspapers starting from the day the original *Nature* article was published in 1993 until 2002. Their analysis revealed that the percentage of newspaper articles on the Mozart effect that mentioned infants increased from 0 in 1995 to about 55 percent of news reports by 1999. Meanwhile, college students – the participants originally studied – were mentioned in 80 percent of articles related to the Mozart effect in 1994 and this steadily decreased to about 30 percent after 2000! Further, media interest in the Mozart effect seemed to increase only after attention shifted from the actual findings to a focus on the putative benefits of music for infants and children. The Mozart effect may be viewed as a 'scientific legend' that drew credibility through association with a scientific study but eventually 'transformed to deviate in essential ways from the understanding of scientists' (Bangerter & Heath, 2004, p. 608). A new story line evolved as media reports frequently overextended the findings to untested populations such as newborns, infants, and children.

The impact of the perceived association between the Mozart effect and infants has been widespread. The findings of Rauscher et al.'s studies are often used for ends not intended by the original researchers, such as the sale of CDs, books, and other products claiming that classical music can 'make babies smarter.' Classical music CDs citing Rauscher and colleagues' studies and claiming various benefits for newborns, expectant mothers, and fathers, are widely available. Articles in the news and popular media with titles such as 'More proof music lifts young IQs' (Ingram, *Toronto Star*, May 1997) continue to reinforce the popular belief that scientists have established a link between music and enhanced intelligence in the developing brain. To date, however, there is no body of scientific evidence that has demonstrated a strong or direction connection between exposure to classical music and enhanced cognitive functioning in infants, and many attempts to replicate the Mozart effect in children have yielded no effect (e.g., Črnčec, Wilson, & Prior, 2006, but see Ivanov & Geake, 2003). Only prolonged active engagement in music such as learning to play a musical instrument – not simply listening to music – has been associated with long-term effects on general intelligence (e.g., Schellenberg, 2006, but see Costa-Giomi, 2004).

That is not to say, however, that music may not have other positive effects on young infants. Music has been effectively used in neonatal intensive care units, especially with high-risk infants. Some studies have shown that playing music to low birth-weight newborns may reduce weight loss and stress-related behaviors, can positively affect oxygen saturation levels and heart rate and respiratory rate, and may reduce the length of hospital stays (e.g., Caine, 1991; Cassidy & Standley, 1995). The Pacifier-Activated-Lullaby method, similar to DeCasper's paradigm described in this chapter, may help preterm infants to more quickly gain self-sufficiency in feeding (e.g, Standley, 1999). This is important as the sucking reflex is often too weak to sustain self-feeding until the fetus is around 35 weeks GA, so that preterm infants must often be fed intravenously. Although not all studies show significant effects of music on physiological measures, the trends are often in a positive direction. Music therapy can also benefit infants of depressed mothers by altering the mother's mood and reducing high arousal levels in their infants (Field, 1998). Singing to infants can also play a role in restoring attachment between infants and family members when bonding has been disrupted by lack of early physical contact with newborns restricted to isolettes (O'Gorman, 2007). The therapeutic use of music with newborns is explored in practical guides such as *Music Therapy for Premature and Newborn Infants* (Nöcker-Ribaupierre, 2004) and *Music Therapy with Premature Infants: Research and Developmental Interventions* (Standley, 2003).

Part Two: Music perception and cognition in the infant

William James (1890, p. 488) used the phrase 'one great blooming buzzing confusion' to describe the world as it might be experienced by the newborn infant. Like many others in his day, he assumed that infants were poorly equipped to process the rich information flooding their senses. This view has changed dramatically with accumulating evidence of infants' acute sensory capacities.

The first part of the chapter described how the cochlea is functionally complete at birth. A few auditory structures are not yet fully developed in the newborn; the tympanic membrane (eardrum) is open at birth and closes by about age 3 years, and the ear canal is straight at birth and assumes the adult S-curve by the end of middle childhood (Snow & McGaha, 2003). The lack of maturity of these structures, however, does not seem to interfere significantly with basic auditory capacities. In the second part of this chapter, we focus on music perception and cognition in the infant. By the end of our discussion, we should have greater insight into why psychologists have replaced the old Jamesian view with the idea of the 'sophisticated infant,' already handily equipped to process complex sensory stimuli.

Orientation to sound

The newborn already responds actively and curiously to sound, as seen in the tendency to turn toward the source of a sound shortly after birth. Spontaneous head-turns towards the sound of a rattle presented to the left or right side have been observed in 2- to 4-day-old newborns (Muir & Field,

Figure 8.1 First measurement of a newborn's head at birth (~13.5 inches, taken directly after birth).

Source: Photograph used with permission from the parents of the newborn. Copyright © Peter Pfordresher.

1979). Some 2-day-old newborns even adjust the position of their heads to track a continuous sound that moves to different points within the same hemi-field (i.e., the same side of the head), though with limited accuracy (Morrongiello, Fenwick, Hillier, & Chance, 1994). After 6 months, most infants can turn their heads within 4 to 6 degrees of a sound source (Morrongiello & Rocca, 1987). And by 7 months, infants can even reach for sound-emitting objects in the dark, relying solely on their sense of hearing in the absence of visual cues (Stack, Muir, Sherriff, & Roman, 1989).

Infants' ability to localize sound does not match adult performance, however. As discussed in chapter 3, our ability to infer where a sound is coming from relies on many cues. Two of the binaural cues are *interaural timing* (if a sound reaches one ear before the other, the source is judged to be closer to that ear) and *interaural intensity* (if the sound is louder in one ear than the other, it must be closer to the source). Infants' poorer performance on sound localization tasks may be due in part to the small size of infants' heads. Figure 8.1 shows the head circumference of a newborn infant, taken a few

Figure 8.2 Infant wearing electrode cap used at the auditory development laboratory at McMaster University.

Source: Copyright © Laurel Trainor and McMaster University.

minutes after birth. Though disproportionately large for the body (newborns' heads are about 1/4 to body mass while the adult ratio is 1/8), the distance between the two ears is much smaller than in adults. Thus, interaural differences in timing and intensity are more difficult to detect. Infants' heads may also not be large enough to produce the 'sound shadow' that conveys cues as to the location of sounds due to sounds having to travel around the head and being absorbed by the head. Further, as young infants spend a lot of time lying on their backs, this too would affect accuracy in detecting sources of sound, as sound localization is more difficult when supine than when standing.

Perception of pitch

The fetus is mainly exposed to low frequencies while *in utero* (for reasons discussed earlier in this chapter) but after birth, infants are more attentive to high-pitched sounds. For both infants and adults, hearing is most acute for frequencies in the 2000 to 4000 Hz range, corresponding roughly to C7 to C8, or the highest C octave on a grand piano (Fernald, 2001). Indeed, for sounds above 4000 Hz, 5-month-old infants perform better at some pitch discrimination tasks than adults, although at lower frequencies infants' hearing is significantly poorer than adults (Olsho, 1984). Thus improvement in frequency discrimination and other areas of listening does not increase in a broad progression across all frequencies; there is significant progress in hearing ability in high frequencies during infancy but low-frequency discrimination advances more slowly and continues into late childhood.

Infants also demonstrate sensitivity to pitch as they imitate voices and other sounds. In an important early study, William Kessen and colleagues trained 3- to 6-month-old infants to sing pitches D, F, and A above the middle C by playing them repeatedly (Kessen, Levine, & Wendrich, 1979). After about 2 weeks of training in the laboratory and at home, most 3- to 6-month-old infants were able to sing back two or three of the pitches with fairly reliable accuracy. Accuracy of pitch-matching was evaluated by five musicians, who judged about 70 percent of the infants' vocalizations to be within a quarter-tone of the target pitch. This finding was surprising at the time, because it was previously believed on the basis of psychoanalytic (Freud, 1924) and learning theory (Bandura & Walters, 1963) that infants could not perform matching behavior until nearly 1 year of age.

There is evidence that sensitivity to consonance and dissonance also emerges at a very early age. Even by 2 months, children prefer harmonic intervals (i.e., pitch intervals formed by simultaneous tones) that are consonant to those that are dissonant (Trainor, Tsang, & Cheung, 2002). This finding makes sense given that the consonance of simultaneous tones likely reflects the mechanics of the cochlea (Plomp & Levelt, 1965, as discussed in chapter 3). However, even sensitivity to consonance in melodic intervals – that is, those formed by *successive* rather than simultaneous tones – emerges early in life, perhaps by 6 months of age (Schellenberg & Trehub, 1996). Other aspects such as pitch perception within a particular scale system, however, may require a little more time and exposure to assimilate. Western adults are much better at detecting tuning changes in melodies written in familiar scales (e.g., major or minor) versus unfamiliar scales (such as Javanese *pélog*, discussed in chapter 15). Young Western infants, on the other hand, discern mistuning equally well in melodies in either scale (Lynch, Eilers, Oller, & Urbano, 1990) until about 12 months of age, after which they show greater sensitivity to tuning changes in melodies based on a major scale, as do Western adults (Lynch & Eilers, 1992).

To date there is little neuroscientific research on music perception in infancy, though this area is growing. One line of inquiry concerns how the brain processes simple pitch differences. Although there is evidence that this perceptual ability seems to be present early based on what we know from behavioral data, it is possible that the brain's response evolves over time. In fact, this does seem to be the case, based on measurements of event-related potentials (ERPs, see chapter 4) in infants. Researchers at McMaster University led by Laurel Trainor have measured EEG responses in infants using a specialized electrode cap (as shown in Figure 8.2) that stays securely on a baby's head during all the squirming and fussing that often occurs. A series of studies by this group measured ERP responses to pitch in 2- and 4-month-old infants, and in adults (He, Hotson & Trainor, 2007, 2009). Participants were presented with single tones that would typically be identical, with an occasional 'oddball' tone thrown in. At issue was how the brain would respond to the 'mismatch' of these oddball tones. It was found that the brains of 4-month-old babies responded very rapidly to oddball tones as do adults. However, the same rapid responses were not found in 2-month-olds – their brains appeared to respond more slowly although they still responded to the mismatch.

Perception of melody

Infants do not simply respond to single pitches in isolation; they also follow the general shape of rising and falling tones within a melody. For instance, Hsing-Wu Chang and Sandra Trehub (1977) showed that 5- and 6-month-old infants know that an exact transposition of a melody they have heard before is not a 'new' melody. The researchers employed the *habituation procedure*, commonly used in infant studies. In this procedure, a sound is repeatedly presented to an infant and a physiological measure (in this case, heart rate) is monitored. The baseline heart rate is first taken. Then an auditory stimulus A is repeatedly presented. Infants' heart rate tends to *decrease* when first attending to a new stimulus, but returns to a normal baseline rate when the infant gets used to the stimulus after repeated presentation. (In infants, a decrease in heart rate seems to indicate interest or attention, whereas an increase often indicates fear.) After the heart rate returns to baseline readings, a new sound (stimulus B) is repeatedly presented. If the infant's heart rate decreases, departing again from baseline level, this suggests that the infant recognizes that it is a new or different sound from before. No change in heart rate would suggest that the infant does not discriminate between the two stimuli.

In Chang and Trehub's study, the same six-note melody was presented repeatedly to infants. Then, either an exact transposition of the original melody (the entire melody shifted up or down by three semitones), or a scrambled melody (the same six tones of the original melody played in a random order) was played. They found that the infants' heart rate did not decrease when the transposed melody was played. However, heart rate decreased when the scrambled melody was played, indicating that the infants detected that it was different from the original melody. The researchers concluded that just as for adults, transposed melodies sound like the original version to 5- and 6-month-old infants, whereas scrambled melodies are heard as 'new' or 'different.'

A 'transposed melody' preserves the melodic contour of the original melody because the exact relationship between each interval is retained, whereas changing the order of the notes for the 'scrambled melody' changes the melodic contour. Chang and Trehub concluded that infants do not attend to the specific pitches of a melody (e.g., A, D, F, etc.), nor to the exact intervals between the pitches, but to the general contour of the melody. This may be why most infants can detect the difference between two six-tone melodies that differ only in one pitch if it changes the melodic contour (Trehub, Thorpe, & Morrongiello, 1985), but often cannot discriminate between two melodies that differ in as many as three or more tones if the melodic contour is preserved (Trehub, Bull, & Thorpe, 1984).

The studies cited above suggest a dominance of relative pitch perception, in that infants treat exact transitions of melodies as if identical to the original melody. At the same time, some evidence suggests that absolute pitch representations may be more prominent than once thought (see chapter 5 for further discussion). Research by Jennifer Saffran and Gregory Griepentrog (2001) led to a surprising conclusion: Infants (8 months old) may learn melodies through absolute pitch rather than through relative pitch. The focus in this paper was on *transitional probabilities* among pitches, which refers to the likelihood that a particular event (e.g., the note C♯) will follow another event (e.g., the note B). Saffran and Griepentrog tested how well infants could learn these transitional probabilities through exposure. They found evidence for this kind of 'statistical' learning, as the infants paid less attention to sequences comprising high transitional probabilities than those for which the transitions were unlikely. The critical twist was that infants could distinguish 'new' from 'old' sequences even when the new and old sequences included the same pitch intervals. By contrast, infants were not as good at distinguishing new from old when the distinction required learning contingencies across pitch intervals (e.g., learning to identify a given melody in multiple 'keys').

How does one reconcile findings such as those of Chang and Trehub (1977) and Saffran and Griepentrog (2001)? One possibility, suggested by the authors of the latter paper, is that most studies focus on infants' ability to extract global information about a melody, but do not test whether infants can learn local information like absolute pitch. Another possibility, suggested by Plantinga and Trainor (2004), is that the answer lies in the kinds of materials that are used. Saffran and Griepentrog used three-note melodies that were atonal in nature. Although infants may not have internalized Western tonality, they have been shown to be better able to learn asymmetric scale structures (scales in which intervals between pitches are not all equal, like the diatonic scale) than the chromatic scale, which is symmetric and used for atonal music (Trehub, Schellenberg, & Kamenetsky, 1999). Consistent with this view, Plantinga and Trainor (2004) found that infants exhibited no preference for a tonal melody in its original key versus a transposed version of that same melody. Plantinga and Trainor, unlike Saffran and Griepentrog, believe that relative pitch dominates infant memory for melody for truly 'musical' sequences. Suffice to say that the study of pitch perception in infancy is a lively one at present.

Perception of rhythm and meter

Just as infants are able to recognize the basic contour of a melody when it is played in another key, they also recognize rhythms when played at different speeds. Sandra Trehub and Leigh Thorpe (1989) found that 7- to 9-month-old infants can differentiate between similar rhythmic patterns (e.g., XX XX versus XXX X, or roughly a five-beat pattern with a silent third or fourth beat) if the *tempo* (speed) at which they are played is changed, as long as the relative durations between each tone is preserved. When the pitches were changed, infants also differentiate between rhythmic patterns that are similar but not the same. The tendency to focus on *relative* pitches between tones of a melody and *relative* durations of tones in rhythms as opposed to absolute pitches and time durations, allows listeners to recognize a song – even if transposed to a different key or played at a different tempo.

Seven-month-old infants also seem to be sensitive to the accent structure of music, as demonstrated in an innovative study by Jessica Phillips-Silver and Laurel Trainor (2005), illustrated in Figure 8.3. In this study, a recording of 2 minutes of a rhythm pattern that had no accented beats was played to 7-month-old infants. As the music played, researchers held the infants in their arms and bounced them either to every second beat, or every third beat, of the rhythm. Therefore, although all infants heard the same unaccented rhythm, the 'felt' accents were different. During the test phase, two versions of the same rhythm pattern the infants had heard before were presented – but this time, the rhythm was clearly accented either on every second or every third beat (i.e., duple or triple structure). Infants who had been bounced in duple time preferred the version that was accented on every second beat, as indicated by longer attentive listening time. Infants who were bounced in triple time preferred the version that was accented on every third beat.

Blindfolding the infants while they were being bounced to the ambiguous rhythmic pattern produced the same results. Again, the infants showed a preference for the rhythm presented in the structure to which they had been bounced before. However, when babies *merely observed* an adult jumping in duple or triple pattern to the ambiguous rhythm while the infant sat still, the infants showed no clear preference for the duple or triple pattern audio track. Thus, visual cues alone do not seem to play an important role in infants' representation of beat structure. The infant has to feel the beat with the movement of *his or her own* body, thus incorporating both auditory and vestibular cues in interpreting rhythmic structure. Similar studies carried out by the same researchers have yielded parallel findings in adults, showing that they too 'hear what the body feels.' Phillips-Silver and Trainor point to a cross-modal interaction between body movement and auditory interpretation of rhythmic structure that emerges early in infancy and is maintained in adulthood: '*how we move* will influence *what we hear*' (2007, p. 535, original emphases).

Figure 8.3 Conditions used in Phillips-Silver and Trainor's (2005) study: (a) Bouncing while watching experimenter. (b) Bouncing while blindfolded. (c) Passively observing experimenter. (d) Preference test.

Source: Copyright © Jessica Phillips-Silver, Laurel Trainor, and McMaster University.

Sensitivity to particular meters may depend on exposure, as shown in a study by Hannon and Trehub (2005a) that was briefly mentioned in chapter 6. Hannon and Trehub found that when Balkan folk melodies of Serbia and Bulgaria were played to North American adults, they could differentiate between changes in the melody that deviated from the original metrical structure versus those that preserved the meter. However, the adults could only do so for Balkan melodies presented in simple meters (which are prevalent in Western music) and not for those in complex meters (which are rare in Western music). North American infants of 6 months, on the other hand, were able to discriminate between meter-disrupting and meter-preserving deviations from the original Balkan folk melody in both simple and complex meters.

Parallel to the finding that infants begin to show greater sensitivity to the tonal systems (e.g., major/minor, *pélog*) of the music to which they are exposed at about 12 months of age (Lynch et al., 1990), Hannon and Trehub (2005b) also reported that at 12 months old, Western infants were only able to perform the task described above for melodies presented in simple meters (and not the complex meters typical of Balkan music). It is possible that infants start with a flexible style of processing tonal and metrical structure, which takes on a more culture-specific character especially after about a year of maturation and exposure to a particular musical repertoire. The *multi-directionality* of development (i.e., the *simultaneous expression of both gain and loss*) is demonstrated here: As infants gain greater specialization and familiarity with a particular musical repertoire, they also lose a certain degree of openness and flexibility in processing sound as music in other ways. This theme of both 'gain and loss' playing a role in becoming specialized will resound in the next chapter on musical development and education.

Memory for music

Thus far, the studies we have discussed have examined infants' immediate responses to musical stimuli shortly after they were presented. But do infants retain music after a period of time has passed? Some years ago, one of your authors (ST) met an 8-month-old infant whose parents claimed that he usually stopped crying at the sound of a particular piece of music: a recording of the pop song 'Y.M.C.A.' Whenever the recording was played while he was distressed, the infant would cry only intermittently, or cease crying altogether. Although his parents tried to soothe the infant with other music, the infant appeared to have a preference for 'Y.M.C.A.' by the Village People. Can 8-month-olds recognize particular songs?

Addressing the question of young infants' memory for music under much more controlled conditions, Jenny Saffran and colleagues asked parents to play the slow movements of two piano sonatas by Mozart to 7-month-old infants once a day for 14 days (Saffran, Loman, & Robertson, 2000). After daily exposure to the music for 2 weeks, parents were instructed not to play the recordings for 2 weeks. The infants were then brought to the lab for testing.

Each infant was tested individually, using an infant research method called the *head turn preferential listening procedure*. In this procedure, a music recording is played when the infant turns his or her head to one side, and switches to another piece of music when the infant turns to the other – much like a 'juke-box' controlled by the infant's head movements. The total amount of time that the infant spends listening attentively to each piece of music is measured, and taken as an indicator of the infant's interest in the music. During the test session, four test items were played to the infants: two 20-second excerpts of the Mozart sonatas to which they had been exposed ('familiar' music) and two 20-second excerpts of new excerpts taken from the same CD recording (the slow movements of two other sonatas by Mozart) for the 'novel' music.

The researchers found that the infants listened significantly longer to the 'novel' Mozart piano sonatas than to the 'familiar' Mozart piano sonatas. In the first part of the chapter, we discussed how newborns are often calmed or soothed by familiar stimuli. However, preferential listening tasks gauge older infants' attention and interest, which is usually oriented to stimuli that are novel. We can infer that the infants in Saffron et al.'s study recognized the 'familiar' music, as they showed significantly less interest to the music that they had been repeatedly exposed to 2 weeks earlier, than to the 'new' music. A control group of infants who had never heard any of the Mozart piano sonatas did not show a significant preference for any of the excerpts.

The CD recordings that the researchers gave to the infants' parents were filled with 10 minutes of Mozart piano sonatas. Thus, the infants seemed to have retained fairly long and complex pieces of music in their memory following a 2-week delay. Further, the infants could not rely on acoustic differences in the music (such as timbre) because the familiar and novel excerpts were all taken from the same CD of Mozart sonatas, all performed on the piano. Subsequent research employing simpler music (folk tunes played on piano or harp as single-line melodies) by Trainor, Wu, and Tsang (2004) and more complex music (piano pieces by Ravel) by Ilari and Polka (2006) have also demonstrated infants' ability to retain music in long-term memory after a delay. Studies on infant memory for music therefore lend some credence to our 'Y.M.C.A.' anecdote!

Infant-directed speech and singing

So far, we have seen that infants demonstrate a sensitivity to elements of music such as pitch, melody, tonality, rhythm, and meter, and appear to be able to recognize familiar music after a time delay. In the final section of this chapter, we discuss how these perceptual and musical abilities also have *social* value for the infant. Our focus is on two auditory stimuli that are often embedded in daily interactions with caregivers: infant-directed speech and singing.

Infant-directed speech

Infant-directed (or ID) speech refers to the distinctive style of high-pitched, melodious speech that speakers often use to communicate with infants. It is distinct from adult-directed (AD) speech with respect to *prosody*, which refers to the 'musical' qualities of speech such as the melody, rhythm, tempo, and intonation of spoken language. Formerly referred to as 'motherese,' the distinctive features of ID speech include higher overall pitch, wider fluctuations in pitch and loudness, slower tempo, shorter phrases containing fewer syllables, and longer pauses compared to AD speech (Cooper & Aslin, 1990).

Pitch contour seems to play a critical role in conveying the message in ID speech. In Anne Fernald's (1989) words, 'The melody carries the message in speech addressed to infants to a much greater extent than in speech addressed to adults' (p. 1505). Common patterns in the 'melody' of ID speech include rising contours for eliciting infant attention or encouraging the infant to vocalize or respond, gradually falling contours to soothe an infant or end a turn-taking sequence, and bell-shaped contours for maintaining the infant's interest or conveying approval (e.g., Papoušek, Papoušek, & Symmes, 1991). These pitch contours are used quite similarly in different languages, including tonal languages (in which pitch inflection conveys meaning, such as Mandarin Chinese) and nontonal languages (such as English and German). Indeed, Mandarin Chinese mothers use pitch contour in speech very similarly to German- and English-speaking mothers. Mandarin-speaking mothers sometimes 'even violate the linguistic tone rules' of the Mandarin language when engaging in ID speech (M. Papoušek, 1996, p. 97, regarding Papoušek et al. 1991).

Many studies show that infants prefer ID speech to AD speech (e.g., Cooper & Aslin, 1990), as indicated by measures of attention such as length of attentive looking time. In fact, infants show a preference for ID speech even if it is not in their native language. For instance, Werker, Pegg, and McLeod (1994) found that infants preferred ID Cantonese speech to AD Cantonese speech – regardless of whether their own native language was Cantonese or English. The authors noted that this finding is 'consistent with the notion that ID communication has universal appeal to infants, over and above any specific differences between languages' (p. 331). One reason for the salience of infant-directed speech is its heightened emotionality, and neuroscientific research has shown that changes in EEG responses correlate with the emotionality of infant-directed speech among 9-month-old infants (Santesso, Schmidt, & Trainor, 2007).

Researchers have observed that ID speech promotes a sort of social synchrony in interactions between caregiver and infant, for example in facilitating joint attention and turn-taking (Adamson & Russell, 1999; Ferrier, 1985). Signals, such as rising contours to initiate dialogue and falling contours to terminate them, may facilitate smooth turn-taking between partners, and thus regulate the timing of events within an interaction. As Custodero (2002) has described it: 'The musical nature of our coordinated conversations with babies is similar to the musical conversations between jazz musicians. Both types of communication require a shared understanding of the musical structure, like "your turn, my turn" ' (p. 6).

Infant-directed singing

While attentive to ID speech, infants show even greater responsiveness to infant-directed singing. When 6-month-old infants are presented with recorded audiovisual presentations of their mothers' ID speech and ID singing, they stop moving and fix their gaze longer at the singing performance (Nakata & Trehub, 2004). Like ID speech, ID singing is characterized as having higher pitch, slower tempo, greater dynamic range and expressive timing, and longer pauses between phrases than non-infant-directed singing (Trehub, Hill, & Kamenetsky, 1997). Fathers also adopt this style of singing to their infants, although they do so less frequently and do not raise their vocal pitch in ID singing as much as mothers do (O'Neill, Trainor, & Trehub, 2001). Even siblings under the age of 3 years also sing at a higher pitch and use an animated, 'smiling' voice when singing to their infant sisters and brothers (Trehub, Unyk, & Henderson, 1994).

Interestingly, even when caregivers are instructed to do their best to sing *as if* they were singing to their infants, infants can discriminate between infant-present and infant-absent singing styles, showing a clear preference for infant-present singing. For instance, Trainor (1996) found that 5- to 7-month-olds looked significantly longer at the loudspeaker that played infant-present recordings of a song than the one that played infant-absent renditions. Mothers' and fathers' singing in the presence of a real infant is also judged by adults to be higher in pitch, slower in tempo, and more engaging in vocal quality ('warm voice,' 'smiling sound') than when doing their best to simulate singing to an infant (Trehub et al., 1997). Facial gestures and body movements such as smiling and swaying (that occur naturally in the presence of an infant) may affect the vocal quality of ID singing. Smiling, for instance, alters the shape of the vocal tract and the acoustic quality of the singing voice to a warmer tone (Sundberg, 1982).

Lullabies and play songs seem to elicit different responses in infants (Rock, Trainor, & Addison, 1999). Six- and 7-month-old infants were presented with recordings of their mothers singing two versions of a song, once in the style of a lullaby and once as a play song. Infants tended to focus on themselves (looking at their own bodies or objects they were touching such as a toy) to songs sung in the style of a lullaby, but directed their attention outwardly by focusing more at the caregiver when listening to the *same* song in a play style. Further, as lullabies and play songs have inherently different musical characteristics (such as differences in pitch range and metrical structure [see Trehub & Schellenberg, 1995]), it is likely that the differences in infants' responses may be more pronounced when singing a song structure that is congruent with the singing style.

The effects of ID singing on sustaining attention, modulating arousal, and coordinating actions of the infant may make singing a useful tool in therapeutic contexts, for instance with infants of depressed mothers and with preterm infants. Depressed mothers often have flat affect in facial and vocal posture – characteristics which do not attract infants' attention for positive mutual engagement. Depressed mothers also often have difficulty recognizing and attending to the emotions of their infants, leading to rather disjointed interactions. De L'Etoile (2006) proposes that ID singing can be combined with other effective approaches such as Field's (1998) 'interactive coaching' by using ID singing as a tool to help depressed mothers attract and sustain the attention of infants, and train mothers to be more emotionally 'in tune' by imitating the responses of the infants while jointly involved in a singing episode. ID singing has also been shown to have physiological effects such as modulation of arousal in infants in the direction of balanced arousal, indicated by changes in salivary cortisol levels after exposure to only 10 minutes of singing (Shenfield, Trehub & Nakata, 2003). Findings such as these have strong implications for therapy with preterm infants, who are frequently under- or over-aroused.

In daily interactions with infants in nonclinical settings, singing is also put to practical use by many parents who sing to infants in the home. ID singing heightens infant attention and reduces body movements (Nakata & Trehub, 2004), facilitating daily routines such as feeding and changing. Singing regulates infants' emotions or arousal levels (Shenfield, Trehub, & Nakata, 2003) which may be helpful in eliciting attention, inviting interaction and play, soothing and reducing crying, and inducing sleep. Like ID speech, ID singing may play an important role in developing caregiver–infant attachment, for example by maintaining infant attention for sustained interaction, promoting emotional synchrony between caregiver and child, and eliciting positive affect in the infant, which in turn can elicit further engagement by the caregiver (Dissanayake, 2000). Singing can contribute to harmony and cohesiveness in the home as it is incorporated into daily routines such as mealtimes, bedtimes, and chores, and to maintain and create new family traditions (Custodero, 2006). In all these ways, infant-directed singing may have adaptive value for both infants and their caregivers.

Coda

This chapter focused on the audition and perception of sound and music, before and after birth. Far from the Jamesian notion of the newborn's experience of the world as 'one great blooming buzzing confusion,' we see an emerging portrait of a competent infant – whose auditory functioning begins a few months before birth. In their first year, they already demonstrate sensitivity to the basic elements of music such as pitch, melody, tonality, rhythm, and meter. Moreover, infants do not only perceive and respond to music; they also create it. In the next chapter, we examine infants' and young children's engagement in musical behaviors – singing, moving to music, and playing with musical toys and instruments – and examine examples of musical inventions and creations of their own making.

9 Early musical development and education

It is often said that 'play is a child's work,' as play provides a rich context for learning essential skills. Far from being a meaningless activity to simply while away the time, developmentalists view play as providing learning opportunities significant to every domain of a child's life: physical, social, and intellectual. In this chapter, we will see how children's early musical improvisations and spontaneous responses to music, and their exploratory play with objects and musical toys, may lay the foundation for musical abilities and knowledge that can be supported and refined through music education.

This chapter opens with an overview of developmental milestones in singing and movement to music in infants and young children. The second section focuses on young children's exploratory play with musical instruments, and the development of children's improvisations and compositions. Finally, we consider how the behaviors and activities that spontaneously emerge in childhood can be refined through scaffolding by adults and by innovative music education methods. Whereas the studies reviewed in preceding chapters focused on experimental studies in laboratory settings where variables can be carefully controlled, we have chosen to focus on more descriptive work and observational studies in naturalistic settings in this chapter. Both approaches have their strengths and limitations, and findings employing a variety of methods can lead to a richer understanding of the complexities of musical development.

Early singing and movement to music

Emergence of song

Infants' vocalizations begin with short and often involuntary bursts of sound. As they develop better vocal control and stronger lungs, their vocalizations become more deliberate and sustained. By around 4 to 6 months, infants often chuckle or coo at the sound of music (Moog, 1976a). By around 6 months, music often elicits a babbling response (Moog, 1976a, p. 59). In language development, 'babbling' refers to vocalizations comprised of repeated consonant and vowel pairings such as 'pababababa.' *Musical babbling* is usually elicited during or immediately after the presentation of music, and is characterized by a wider variety of syllables and greater variation in pitch than speech babbling. The pitch is usually centered (very approximately) on the F above the middle C and spans about one octave (Moog, 1976b, p. 40), and the melodic contour is typically descending, with upward leaps between phrases (Davidson, 1994; Moog, 1976a, p. 60).

Musical babbling does not yet reflect the characteristics of the musical repertoire to which the infant is exposed. The pitches do not adhere to a specific tonal scheme; for instance, pitches in the musical babbling of Western infants are microtonal or fall between the musical tones of the diatonic scale (Dowling, 1984; Moog, 1976b, p. 40). Musical babbling also tends to be 'rhythmically amorphous' and does not yet conform to a regular pulse (Moog, 1976a, p. 60).

By around 2 or 3 years, children begin to engage in *spontaneous singing*, or improvisations of their own that are not yet attempts at conventional songs (Moog, 1976a, p. 100). Figure 9.1 shows an example (taken from our own archives) of 2-year-old Freya's vocal improvisations while playing with a doll,

Figure 9.1 Spontaneous song by 2-year-old Freya. The broken stave represents her tendency to sing in phrases rather than metrically, and parentheses denote speech-like tones.

Source: Recording obtained by Amanda McFarland and transcribed by Siu-Lan Tan and Peter Pfordresher, used with permission of the child's parents. Copyright © Siu-Lan Tan and Peter Pfordresher.

illustrating many typical features of spontaneous songs. As is common of spontaneous songs by children of her age, Freya's song is constructed out of short phrases repeated many times with little development or extension of ideas (Dowling, 1984). Every phrase was sung in one breath with the same general contour – rising and falling with the natural arc of a breath – except for the ending. Our notation is approximate as many pitches were microtonal and the endings of phrases switched to a speech-like tone. There was no clear feeling of meter as each phrase was sung in one breath with frequent pauses between breaths. Her 'performance' was expressive and accompanied by actions, 'rolling' (rocking) the doll and standing up and singing loudly at the 'stand up!' lyric. She already has the idea of a graceful song ending: the last phrase is the only one that features a rising contour and a sustained (rather than clipped) last tone.

While the general structure of the song bears some similarity to conventional songs (e.g., melodic and textual repetition, and an ending), spontaneous songs of 2-year-olds like Freya do not yet reflect the finer characteristics of a particular musical repertoire such as the tonal system, or preferred meters. However, as Figure 9.2 shows, the spontaneous songs of 4- to 5-year-olds already begin to incorporate some general features of familiar musical genres to which they are exposed, such as 'pop' songs. The first two song excerpts in this figure were drawn from spontaneous songs sung by researcher David Hargreaves' son, Jon, and the last two excerpts were sung by

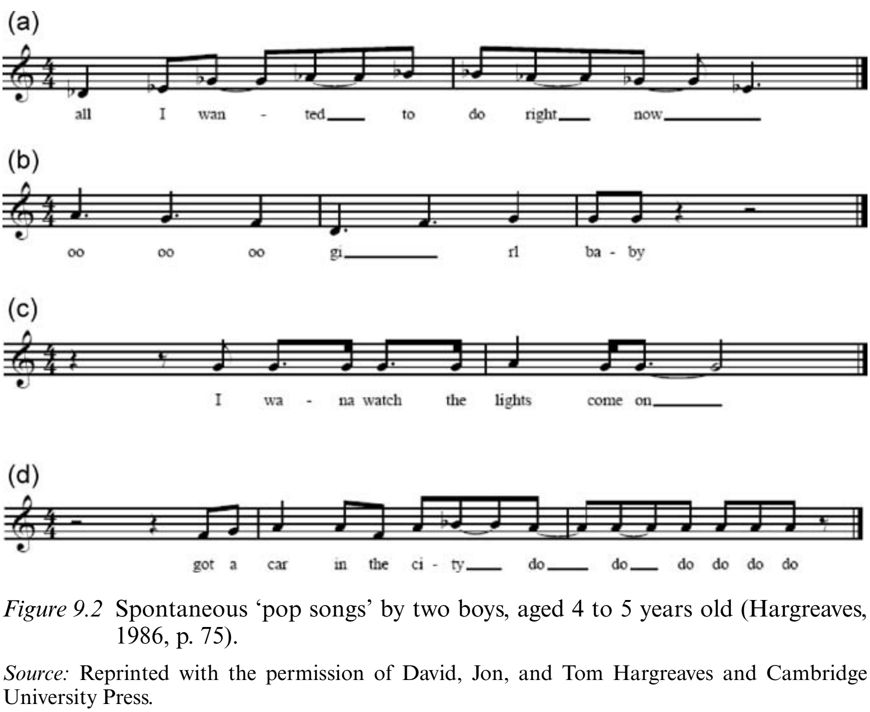


Figure 9.2 Spontaneous 'pop songs' by two boys, aged 4 to 5 years old (Hargreaves, 1986, p. 75).

Source: Reprinted with the permission of David, Jon, and Tom Hargreaves and Cambridge University Press.

his brother Tom in a spontaneous 'recording session' at home involving both boys on vocals, percussion, and piano at age 4 to 5 years (D. Hargreaves, personal communication). The influence of 'pop' songs on Jon and Tom's singing is reflected in their use of syncopation and in their chosen lyrics, and the musical ideas are longer than Freya's at age 2 years.

Melody is not always the most prominent feature of children's early improvised songs. Words and rhythm are more distinctive in children's *chanting*, emerging during early childhood. The focal tone of children's chant falls within the range of the D♯ to F♯ above the middle C (Moorhead & Pond, 1941, p. 19), and the chant generally follows a falling contour with the most common interval being a minor third. In a naturalistic study of children's chant, Moorhead and Pond observed that children often chant during solitary or social play, when engaged in repetitive motor activities such as running and jumping, and when giving commentaries of events (e.g., announcing a sudden outbreak of rain). Onomatopoeia and words created to imitate sound ('the rhumba, the rhumba, ba, ba, ba' to mimic the sound of a wagon [p. 13]) were often incorporated into the verse. Nonsense rhymes were also common ('a pumpker, a crumpker. A pumpker, a crumpker.' [p. 14]). Some episodes of solitary chant were observed but children often improvised together, demonstrating a surprising degree of social organization:

A: 'Red, red, gingerbread.'
B: 'Red, red, go to bed/Red, red, go to bed.'
D: 'Stand on your head/With gingerbread.' (p. 14)

Along with songs and chants of their own making, Moog observed the emergence of what he called 'potpourri songs' by the ages of 2 or 3 years. *Potpourri songs* combine spontaneous singing with fragments of words, rhythms, or melodies of conventional songs (1976a, p. 100). An example of a potpourri song sung by 2-year-old Freya is shown in Figure 9.3. During snack time, she was proudly demonstrating her newly acquired skill of eating with a fork. After murmuring syllables such as 'why, why, why' she began to sing the phrase 'Mama come here watch Rah (Freya) do it' to the tune of 'London Bridge is Falling Down,' with some variations of her own (resembling an introduction and coda). When reproducing 'London Bridge' she used a consistently song-like voice that was very different in quality from her spontaneous song, and stayed remarkably close to the right pitches (except for singing a tone closer to a C than a D for the beginning of 'my pfair lady').

Singing to imitate a model is achieved by about age 3 years (Moog, 1976a, Figure 8, p. 99). When songs were sung to 1-year-olds, very few (only 6 percent) were able to give rough imitations of song fragments (p. 77), and by about age 2 years most children could only reproduce a general outline or rough impression of the song. The children tended to imitate words first, then rhythm, and finally melodic contour. By age 3 years, Moog found that most of the children could imitate a whole song although not very accurately,

Figure 9.3 ‘Potpourri’ song by 2-year-old Freya. No time signature is included as roughly equal emphasis was placed on each note; bar lines have only been added for convenience of reading.

Source: Recording obtained by Amanda McFarland and transcribed by Siu-Lan Tan, used with permission of the child’s parents.

particularly with respect to pitch accuracy (1976a, p. 99). Words were still the most accurately reproduced feature, and rhythmic errors were often made if the rhythm did not fit the syllabic pattern of the lyrics. Freya was quite accurate in her singing of the pitches in the ‘London Bridge’ melody, but the structure of this song is particularly simple and supplying her own (repeated) lyrics may have facilitated accurate reproduction of the tune. When singing other nursery rhymes at age 2, her renditions were often rather fragmented and incomplete. By the age of 4 or 5 years, most children know a small repertoire of conventional songs, though pitches and rhythms are not yet precise. The sense of tonality is not yet stable so children often shift into different keys between phrases until 5 years or older (e.g., Shuter-Dyson & Gabriel, 1981).

By age 5 years, many of the basic skills of song-singing are in place: although each interval may not always be precise, children’s singing reflects a clearer sense of tonal center across phrase boundaries, an underlying pulse is apparent, and the ability to sing with expression is acquired (Davidson, McKernon, & Gardner, 1981; Shuter-Dyson & Gabriel, 1981). For instance, when asked to sing songs to make an experimenter happy or sad, girls and boys as young as 4 years old sang the happy versions faster, louder, and at a higher pitch than sad versions (Adachi & Trehub, 1998). They were able to sing in this expressive manner when reproducing familiar nursery rhymes, and also while creating their own melodies and lyrics in happy and sad songs that they improvised. Children’s invented songs also increasingly convey conventional features of the cultural repertoire, such as repetition of ideas, four-phrase form, and a sense of closure in Western 5- to 7-year-old children’s vocal improvisations (Davies, 1994).

Singing in tune is a difficult task as it requires the complex coordination of internal feedback mechanisms and auditory feedback, such as the ability to adjust the posture of the vocal cords to the sound of one's voice (see chapter 11). When learning songs they are taught, young children typically begin with a chant-like approach that focuses on the lyrics, then sing within a narrow pitch range resembling the speaking range, and then move through a phase in which they alternate between speaking and singing range before finally consistently using a wider singing range (Rutkowski, 1997). Large intervals continue to pose a challenge to young children as well as adults. For instance, both children and musically untrained adults often compress the octave leap in the third phrase of the 'Happy Birthday' song (Davidson, 1994). Providing children with a vocal model (especially a voice similar to their own) increases accurate pitch-matching in singing instruction of elementary school children (Green, 1990). Also, providing children with a real-time visual display of their own voices (e.g., using computer software to graphically show the target pitch and the pitch the child is singing) appears to be helpful in pitch-matching, and has been shown to be more effective than providing verbal feedback after singing (Welch, Howard, & Rush, 1989).

The potpourri song and learning of conventional songs marks the transition from a focus on improvisational and creative aspects of song to the acquisition of a particular musical repertoire. Here we see the *multidirectional* nature of development, as *both gain and loss are jointly expressed* in the development of singing: The child loses a measure of flexibility and spontaneity in their song production but, in assimilating the values and conventions of a specific repertoire, becomes a more specialized member of a particular musical community. We also see a shift from solitary singing to more collaborative creations, and enculturation into group ensemble singing. With the ability to sing a common repertoire of learned songs comes the possibility of singing in synchrony with groups so that singing, an action once done perhaps for oneself, takes on an increasingly social character.

Movement to music

While the emergence of singing is a relatively well-researched topic, much less is known about the development of movement to music. As described in the previous chapter, infants turn their heads and show an alerting response to sounds, especially high pitches and the human voice in melodious speech and singing. With the refinement of voluntary motor skills, they respond with more varied and coordinated movements of the body. From 4 to 6 months, many infants respond to music by turning toward the sound and ceasing movement as if listening with rapt attention. After about 6 months music begins to elicit increased movement, accompanied by musical babbling and other vocalizations (Moog, 1976b). As most infants begin to sit up at around this age, their responses are more noticeable as the upper body is unconstrained. Movements to music often emerge first in the head and neck in the form of nodding motions to rhythmic music, often accompanied by vocalizations. When the infant can sit up securely and proprioceptive skills are advanced enough to make postural adjustments (around 7 months), the torso is also engaged. The infant begins to bounce, or sway the trunk forwards or sideways. The arms are also free to swing when sitting securely, so that many infants appear to make rough 'conducting' motions to music (though only for short bursts, and not yet synchronized with the music).

As infants begin to stand (about 11 months) and walk (just before 12 months), they engage more of the body in dance, bending the knees to music and stepping with alternating feet. Approaching age 2 years, increasing strength and control of the feet and toes leads to the ability to tiptoe to music. As coordination of upper and lower limbs improves, conducting movements with the arms are seen more frequently than before, and are combined with knee-bends or stepping (Moog, 1976a). With increased mobility and balance, toddlers begin to use more space in their movements – running, turning, and spinning to music.

The development of infants' movement to music reflects two basic trends in physical and motor development. There is the *cephalocaudal trend*, or the tendency for muscular strength and control to proceed from the top of the body and progress downwards (beginning with bobbing the head to music and moving downward to the engagement of the trunk, legs, feet and toes), and the *proximodistal trend*, or the tendency for muscular strength and control to progress from the core of the body outward to the extremities (swaying the trunk before engaging the limbs, fingers, and toes in movement to music). While change occurs rapidly in the infant and specific milestones may appear to be unrelated or random when singly observed, it is interesting to observe that, as a whole, motor development and motor responses to music progress in an orderly fashion.

At about 18 months, another milestone is achieved as many toddlers begin to dance with a partner (Moog, 1976a). At first, toddlers' movements are not yet coordinated with those of the partner. Only gradually, the emphasis shifts first to coordinating actions with others and finally progresses to matching movements to the music. In particular, the ability to move in synchrony with music takes a long time to master and may not be achieved until well into middle childhood (about 8 or 9 years). For instance, Trehub (1993) found that the rousing music of *McNamara's Band* elicited more bouncing movements in 9- to 13-month-olds than Schumann's slow and lyrical *Traumerei*. However, these motions captured little more than the general pace of the music, and were not yet 'in time' with the music. Most of the bouncing, swaying, and dance-like movements of the 2- or 3-year-old are also not yet synchronized with the beat of the music (Moog, 1976a).

Even simply beating in time to music is difficult for young children. Malbrán (2000) found that 3-year-old children have difficulty keeping a regular beat when asked to tap the pulse of short pieces of music, even though the music was in strict march time played at a moderate speed. Further, the ability to engage the whole body in musical movement lags far behind the ability to move isolated parts of the body rhythmically. Rainbow (1981) observed over 150 preschool-aged children during a 3-year longitudinal study on the development of rhythmic abilities. He found that less than 4 percent of the 3-year-olds could clap hands and march simultaneously in time to music, and less than 15 percent of the 4-year-olds could perform this task. The 3- and 4-year-old children were more accurate when either clapping or tapping a steady beat to music than when marching to the beat of music, and least accurate when marching and clapping simultaneously to music.

One important limitation of children's ability to synchronize is the range of *tempi* (or speed of music) to which they can adapt. Children produce much faster spontaneous *tempi* than do adults (e.g., Drake, Jones, & Baruch, 2000; see chapter 6 for further discussion of spontaneous tempo). Drake and colleagues further found that 4-year-old children tapped to a much faster 'beat' when allowed to 'find their own beat' than did older children and adults, and were less flexible in the range of speeds to which they could synchronize. In her observation of young children's responses to recorded music in an Israeli nursery, for instance, Gorali-Turel (1999) found that children often attempted to imitate adult models' movements to music, but only if the adults incorporated the children's natural *tempo* and style of movements.

In the early development of movement to music, we see a progression from short bursts of physical activity to more sustained and controlled motor actions, from the infant's own improvised movements that capture only the broadest acoustic features of the music, to children's ability to match movements to music and coordinate one's actions with others. As with the emergence and development of singing, children gradually become socialized into ways of interacting with music reflecting the values and style of a particular musical culture.

As we shall see, innovative approaches to early childhood music education are attuned to the ways in which young children naturally integrate music into their daily lives, and encourage them to initiate and shape their own learning activities before more structured music instruction begins. Consistent with the current literature in music education (e.g., Smithrim & Upitis, 2007), our discussion of music learning in young children is broadly framed by the 'constructivist' theories of Piaget and Vygotsky. In both views, the child is driven by an inherent curiosity and actively *constructs* their own knowledge, but the pathway to mastery takes somewhat different routes.

Musical play and invention

Exploration of sound

The active role of the learner is fundamental to Jean Piaget's (1896–1980) theory of cognitive development. His constructivist approach assumes that learning is driven by one's own curiosity and active exploration. In Piaget's (1936/1952) view, the *sensorimotor stage* (birth through 2 years) is a period of active exploration through the use of senses and motor actions. These explorations, at first isolated and sporadic, become gradually more organized. Sensorimotor explorations of sound may occur through *primary circular reactions*, or repeated actions involving the infant's own body (Piaget, 1936/1952). As the infant explores a range of vocalizations during vocal play, she repeats those that bring about interesting sensory effects (such as popping sounds made by opening and closing the lips, and clapping the hands together). Actions that may have arisen spontaneously through vocal play become increasingly deliberate and sustained, as cooing and babbling and more intentional motor movements emerge.

Further exploration leads to the discovery of the consequences of repeated motor actions and manipulation of objects. An infant who repeatedly hits a metal spoon against a table (a *secondary circular reaction*, or repeated action oriented to the external world) or throws a series of different objects on the floor (a *tertiary circular reaction* or action that is repeated with variations, bringing about different outcomes) quickly discovers that physical actions on objects produce different timbres, pitches, and rhythms. As you may have noticed, infants do not sit still for long. They constantly grab objects, explore them by mouth, shake them noisily and hit them against surfaces, throw them on the ground, and repeat the cycle over again. Like a scientist methodically repeating experiments, infants use their senses and motor actions to build a repertoire of knowledge about the physical properties of objects in their environment.

In an interesting demonstration of infants' ability to match sight and sound in musical instruments, Anne Pick and her colleagues found that 7- to 9-month-old infants can see a musical instrument that is being played and match it with the sound of the instrument (Pick, Gross, Heinrichs, Love, & Palmer, 1994). The study employed a procedure developed by Spelke (1976), in which infants see two videos presented side by side on two screens, but hear only one soundtrack that matches one of the videos (see Figure 9.4). In Pick et al.'s study, the two videos each showed a musician playing a different musical instrument, but each performing the same song in synchrony. A loudspeaker (located between the two video screens) played the matching soundtrack to one of the musical instruments that was being played. Some instruments included in the study were flute, clarinet, trumpet, trombone, viola, and cello. By 7 to 9 months, infants looked significantly longer at the screen showing the musical instrument that matched the soundtrack than at the other screen.

How are such young infants able to do this? Recall that the musicians were all playing the same song in synchrony, so the specific melody and rhythm of the piece being played did not serve as cues for matching sight and sound. Pick and colleagues interpreted their results in the context of the 'ecological' approach (Gibson, 1979), which differs in important respects from the Piagetian tradition. Nevertheless, certain aspects of their results resonate with the Piagetian framework. Through repeated physical manipulation of

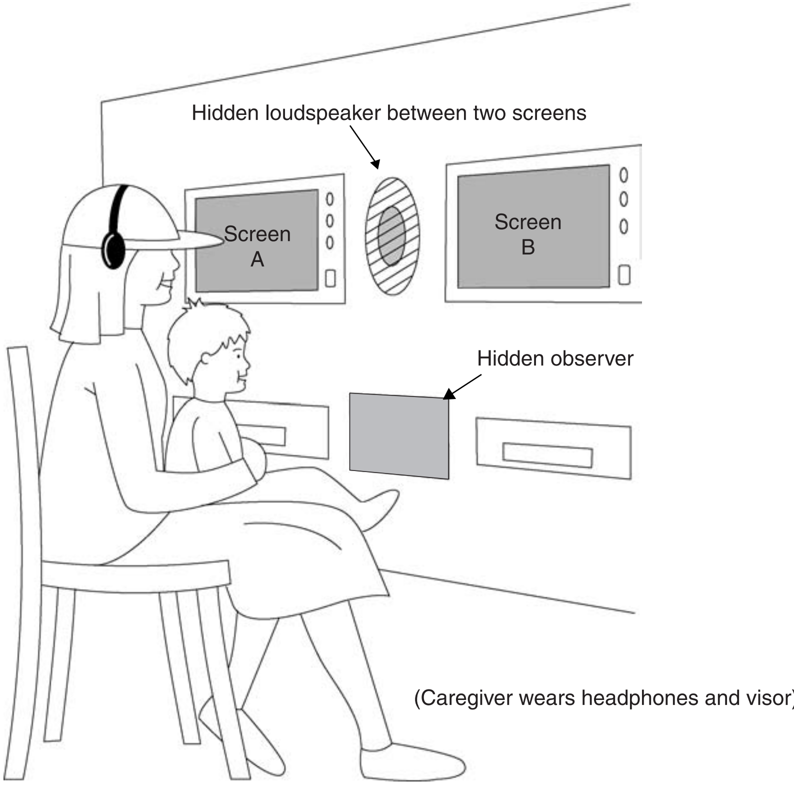


Figure 9.4 Apparatus for studying intermodal (visual-auditory) perception. Note that the parent wears a visor and headphones so only the infant can see the screen and hear the soundtracks.

objects in their environments, infants make discoveries about the sound-making possibilities of objects. With a combination of active experience and biological maturation, these sensorimotor schemes may become further refined so that infants form ideas of the sound that will be produced by an object of a certain shape or size, or by particular actions performed on it.

Just as the infant discovers the sound-making possibilities of a spoon or tin cup by manipulating it, the young child in the *preoperational stage* (2 to 6 years) constructs knowledge through more complex forms of active exploration. The discussion that follows focuses on musical exploration during early childhood. As we have already outlined the emergence of song and movement, we shift our focus to children's musical play with toys and instruments. Two themes in the literature on early childhood music education are the importance of self-initiated musical play, and the need to focus on process (not just product) in young children's early music-making.

Musical play

A theme in both the practice and research in music education concerns the role of young children's self-initiated or 'free musical play' in music learning. In many traditional teaching approaches, the teacher directs and structures most learning activities. In free musical play, however, the teacher provides rich opportunities for musical exploration and lets the child structure their own learning through exploration, while guiding and supporting their discoveries. Such a pedagogical approach may be seen as an application of Piaget's theory to practical educational contexts, as the idea of learning through active exploration and discovery is at its core.

In an early study of unstructured musical exploration, Gladys Moorhead and her colleagues observed 2- to 6-year-old children's free play with musical instruments at their Pillsbury Foundation School in California (Moorhead & Pond, 1942; Moorhead, Sandvik, & Wight, 1951). As a central goal of the school's curriculum was to allow children to discover 'natural forms of musical expression,' emphasis was placed on children's freedom to explore musical instruments with minimal adult intervention. No formal music period for instrumental study was assigned, and children were free to independently explore a variety of musical instruments from around the world (such as gongs, cymbals, bamboo sticks, *maracas*, drums, guitars, violins, and piano) at any time. Consistent with Piaget's observations, young children's explorations of musical instruments have a repetitive and methodical character, as if 'investigating' the possibilities of sound for themselves. When left to their own devices, common patterns were observed in the way children explore sound in novel instruments. For instance, when initially exploring a new instrument, most children appeared to be drawn first to the *timbre* of an instrument:

> *April 2.* Playing the Chinese theater drum with a hard-headed striker, Carl [3 years] discovered that the drum produced two different tones, one from the central stretched skin, one from the wide skin-covered wooden frame. He played a fast staccato beat, fairly steady, intent upon the effect produced by alternating the two tones. He then used the French bell series, alternating between the bells and the theater drum. Apparently disliking the harsh sound of the bells when struck by the hard striker he was using, he struck them seldom, continuing his interest in the two drum tones.
>
> (Moorhead et al., 1951, p. 95. See also Moorhead & Pond, 1942, pp. 42, 44–45)

Also emerging early during initial exploration was experimentation with *loudness*, especially in testing the softest and loudest sounds an instrument can make. Experimentation with variations in *tempo* (speed) and *rhythm* emerged later, only after the child was already accustomed to the instrument.

When first playing new instruments, children tended to adopt a fairly regular and unaccented beat, as seen in Carl's exploration of the drums and bells. After thoroughly exploring timbre and volume, children often began to experiment with different rhythmic patterns and changes in speed, as did Roy:

> *May 12.* Roy [4 years] hung a small Indian tom-tom around his neck and walked around with regular steps, drumming with two sticks in simultaneous beat, the beat coinciding with the right foot; almost without accent. He now began to drum in three-eighths time, using two sticks simultaneously . . . He accelerated his tempo continuing for a time the effort to make his feet and drum beat coincide . . .
>
> (Moorhead et al., 1951, p. 98)

Just as the infant explores sound through active exploration of objects, the child has much to discover about music through playful experimentation with musical toys and instruments: 'In their own way children are analyzing sound. They are inventing and discovering different timbres; they are learning to identify, compare, discriminate, classify, arrange and rearrange sound' (Shelley, 1981, p. 26). Perhaps most importantly, free musical play allows children the time to become fully engrossed in a self-directed activity. In contrast, traditional teacher-led classroom activities are often changed every few minutes in order to hold the interest of the class, providing few opportunities for deep absorption in an independent activity. Under these conditions, exploratory play is often prematurely terminated (Berger & Cooper, 2003), fracturing the child's deep concentration and disrupting the process of learning and discovery.

Observing process

Tarnowski (1999) points out that like any other form of play, musical play 'is an activity in which being engaged in a *process*, rather than achieving a final product, is the goal' (p. 27, emphases added). To the child, the real value is in the *doing*, not in the final outcome (see Figure 9.5). However, adults often do not consider this type of exploration to be musical nor recognize its role in music learning (Fox, 1989). Few adults would regard actions such as scrubbing the keys of a toy xylophone with the wrong end of the mallet as instructive to musical learning. Such activities are likely to be regarded as mere annoyances, as adults are often impatient to hear more 'mature' products of music-making:

> Adults are accustomed to recognize musical production of children according to arbitrary standards of their own which merely draw lines near the peak of the enormous body of their musical experience and production. Each adult draws this line to suit himself, rejecting most of what is real music for the child at his own level.
>
> (Moorhead & Pond, 1942, p. 32)

Figure 9.5 A toddler's first attempt at music performance at age 14 months.

Source: Photograph used with permission from the parents of the child. Copyright © Peter Pfordresher.

Researchers, too, tend to focus on the product of children's music-making, often without much attention to process. Children's spontaneous vocalizations or improvisations with instruments are typically recorded, transcribed, and analyzed in research studies. However, young children's improvisations do not naturally fit the pitches, rhythms, or text of adult forms of notation; transcribing the sounds may therefore impose a structure or organization that is not found in the original. Further, much richness is lost when children's improvisations on musical instruments are analyzed without examining how the sounds are made, or considering the whole context in which the vocal play or instrumental play took place: 'the method of transcribing audio-recordings [of young children's improvisations] into notation *lifts off the sonic layer of the music* and neglects the contextual detail: the child's bodily movements in interaction with the instruments and with others around, and information about the setting' (Young, 2003a, p. 176, emphasis added).

In particular, a product-oriented approach has resulted in a lack of attention to the *kinaesthetic* aspects of young children's music-making, such as the role of body movements in generating sounds from instruments. Sound and action are often intrinsically interlinked for the young child. As Susan Young (2003a) has described it: 'Considered from the child's perspective, the sounding result is inseparable from bodily involvement. The sounding ideas are perceived as analogues of their movement, one facet of an inter-sensory whole' (p. 56). This was clearly demonstrated in Young's observations of nursery school-aged children playing with xylophones during spontaneous play. The children's musical explorations progressed in an organized manner through a series of exploratory movements. The first type of action the children typically employed was to strike the keys vertically. Next, children often experimented with various horizontal sweeps – swishing in one direction first, then zigzagging in two directions, and finally moving on to circular motions. They tended to restrict left and right hands to separate areas of the xylophone at first, before experimenting with crossing the dominant hand over the other hand, and finally crossing both hands. Children also appeared to learn about dynamic range through connecting specific actions (such as fast versus slow actions, disjointed versus flowing movements) with their expressive consequences.

Viewing child development more broadly, we often see connections between milestones in different domains. For instance, it is interesting to note similarities between Young's description of children's initial play with xylophones and John Matthews' (1999) observations of infants' scribbles. Matthews observed that infants typically make their first scribbles with vertical motions (vertical arcs) followed by horizontal sweeps in one direction (horizontal arcs), zigzags in two directions (push-pull actions) and finally circular scribbles (scribble whirls) – exactly the same sequence as was observed in young children's first explorations and improvisations on xylophones. Observing the process and not simply the products of these actions reveals the underlying similarities in the way children explore 'space' using different materials, and points to the orderly nature of the progression of motor and cognitive development.

Children's musical compositions, too, can be more richly studied by examining both process and product. John Kratus (1989) asked children aged 7, 9, and 11 years to compose on an electronic keyboard for 10 minutes. The 7-year-old children spent most of the allotted 10 minutes engaged in 'exploration' (playing new tone combinations and rhythms without repeating ideas). The 9-year-olds engaged in exploration, but also spent half the time engaged in 'repetition' and 'development' (variations of what they had played earlier in the same session). The oldest group, the 11-year-olds, divided the 10 minutes evenly between exploration, repetition, and development. Consequently the youngest group's strategy for composing was similar to experimentation or improvisation, while the older children's compositions showed the beginning of basic form.

The shift from process to product

When do children shift their attention from the process of music-making to the products of their musical efforts? In his observations of children's compositions, Kratus (1989) concluded that the 7-year-olds were process-oriented in their approach to composition, as 'the process of exploring new sounds took precedence over the creation of a single composed product' (p. 19). The 11-year-olds, on the other hand, were more product-oriented and 'seemed to be more focused on the task of creating a single product than on exploring new sounds' (p. 19).

Kratus' study employed a *cross-sectional* design (comparing the development of different children, each of different ages to infer the trajectory of development). Using a *longitudinal* design, Brophy (2005) followed the *same* 62 children for 3 years, from the age of 7 to 9 years. He also observed a shift toward increased musical organization (e.g., metric organization, repeated rhythmic motifs) in their instrumental improvisations, occurring most noticeably between ages 7 and 8 years. Brophy concluded that 'process orientation may be ending for most children around age 7, and product orientation may begin around age 8' (p. 131). Interestingly, we see a similar shift in other domains of development during middle childhood. For instance, Luquet (1913, and others in subsequent research) observed that by around 7 or 8 years, children become preoccupied with conveying realism in their drawings, and begin to compare the images they create to an external standard.

Spiral model of musical development

Kratus' and Brophy's studies provide a glimpse into the developmental changes occurring in children's musical creations during middle to late childhood. A more comprehensive approach to characterizing musical development was undertaken by Keith Swanwick and June Tillman (1986), who collected and analyzed 745 improvisations and compositions of children between the ages of 3 and 11 years old in a 4-year longitudinal study. Their findings suggested an orderly sequence of musical development, which led to the formulation of the model shown in Figure 9.6.

The model takes a spiral form in order to represent the alternation between two polarities, loosely based on Piaget's ideas of assimilation and accommodation: the modes on the left side of the spiral represent the ability 'to relate experiential data to our internal systems of meaning' (assimilation) while the right side of the spiral represents the ability 'to modify these systems when they cease to be adequate to interpret experience and sustain coherence' (accommodation) (Swanwick, 1994, p. 87). The initial studies were conducted with British children in the United Kingdom, but Swanwick (1994) also found evidence for similar trends in musical development of children in other countries. What follows is a brief summary of the spiral model, and the eight levels which each build onto those previous to it. (Approximate ages were provided in the original 1986 model, but Swanwick [1994] believes that too much emphasis has been placed on the ages, and does not provide ages for each level in later work.)

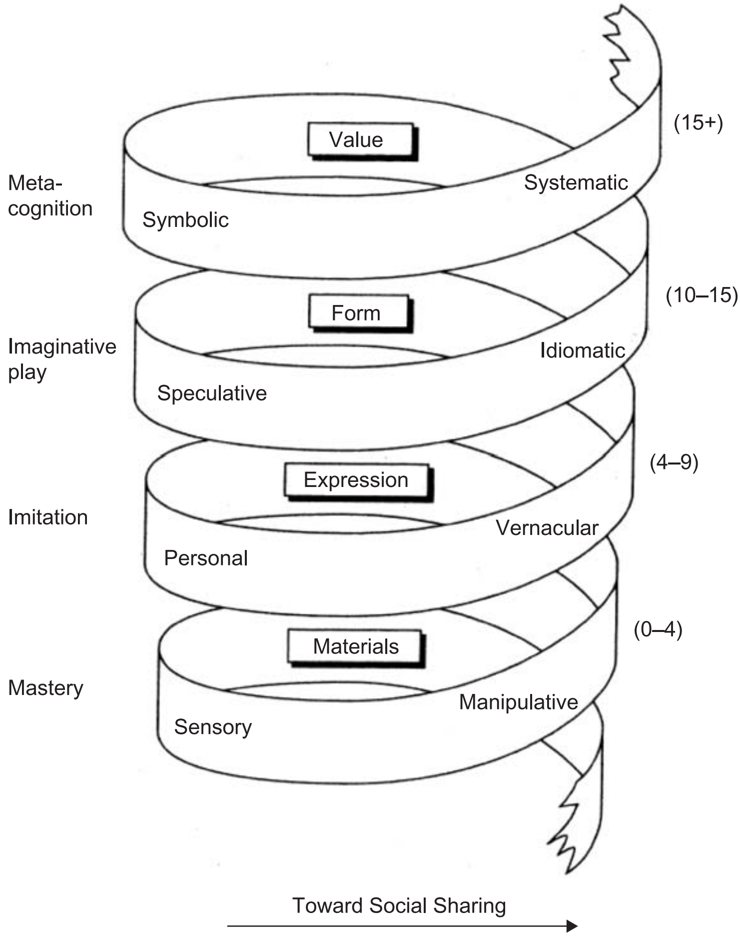


Figure 9.6 Spiral model of musical development (Swanwick & Tillman, 1986, p. 306).
Source: Reprinted with permission of Cambridge University Press.

Materials

Level 1 – Sensory (up to about age 3 years): The child is concerned with sound, especially timbre and dynamic range. Compositions are not yet organized with respect to expressive or structural significance, as the emphasis is on exploration of sound. The first level of the model is consistent with Moorhead and Pond's (1942) observation that children are initially attracted to timbre and dynamic range when first exploring instruments.

Level 2 – Manipulative (4- and 5-year-old children): Building onto the previous level, fascination with sound shifts to a focus on learning to manipulate and control the materials to make sound. The child becomes increasingly skillful at handling instruments, producing sounds such as *glissandi*, trills, and tremolos. Compositions are more sustained but still unstructured, and compositional structure is often determined by the physical characteristics of the instruments. Earlier in this chapter, for instance, in Young's (2003a) observations discussed earlier, the physical layout of the xylophone may have structured the children's initial explorations and improvisations. Rather than being guided by any internal representation of form or expressive character, musical compositions at this level may be structured by the affordances of the particular instrument.

Expression

Level 3 – Personal Expressiveness (no age provided): As this model is cumulative, each level builds onto the previous ones. Here, the focus shifts from manipulation of musical materials to an interest in the expressive possibilities, such as dynamic level and speed. Therefore, the performance may have dramatic character and climax, but without much attention to overall structure (i.e., little deliberate repetition or development of ideas). This level is consistent with Moorhead and Pond's (1942) observation that following an initial attraction to timbre, children begin to experiment with volume (loud and soft, *crescendo* and *decrescendo*) and *tempi* (such as fast and slow, *accelerando* and *decelerando*).

Level 4 – Vernacular (emerges at about age 5 to 6 years, but not clearly established until about 8 to 9 years): Repeated rhythmic and melodic motifs begin to emerge, and the improvisations or compositions begin to show the early beginnings of form. This is consistent with Kratus' (1989) observation of how children gradually employ less exploration of novel ideas, and increasingly more repetition and development of previous ideas, in their compositions between age 7 and 11 years.

At this juncture, we see a shift from a focus on the *expressive* character of music to its *formal* aspects, and from process to product. The last four levels, labeled as 'Form' (level 5 and 6) and 'Value' (level 7 and 8), are associated with late childhood and beyond, and are only briefly summarized here. During *Level 5 – Speculative* (10 years), form is further refined. In addition to repetition, there is also experimentation with deviation and elements of surprise. During *Level 6 – Idiomatic* (13 or 14 years), form is further integrated into a recognizable style such as classical, jazz, or popular music. *Level 7 – Symbolic* (unlikely before age 15 years) is characterized by a consciousness of stylistic idiom as the individual identifies strongly with certain pieces of music, and becomes personally committed to certain musical values over others. Finally, during *Level 8 – Systematic* (no age provided), one is not only able to understand principles of musical style, but may go beyond them to create novel systems of music. The broken ends of the spiral represent the continuous reactivation of the sequence as the model is recursive. When coming upon musical genres or instruments that are new, for instance, a person may again move back through exploration of materials and expressive possibilities to the higher levels.

Children's earliest improvisations are a form of musical play, initially engaged in for the sake of the activity, without a focus on assessing the product. Once again, in the progression of the spiral model – as in many other domains – we see *gain and loss* jointly expressed in development. Older children may lose some of the spontaneity and inventiveness of early childhood but they are moving towards more organized and intentional creations.

Supporting early musical development

Learning in social context

While Piaget viewed children as independent explorers who construct knowledge from discoveries made largely on their own, Lev Semionovich Vygotsky (1896–1934) was more concerned with the social and cultural influences on children's cognitive growth in the rich context of daily social interactions with parents, teachers, and more competent peers. Vygotsky believed that 'learning awakens a variety of internal developmental processes that are able to operate only when the child is interacting with people in [the] environment and in cooperation with . . . peers' (1978, p. 90). In his view, children are born with elementary mental functions such as attention, sensation, perception, and memory, but it is through interactions with symbolic representatives of their culture that more sophisticated higher mental functions such as logical and abstract thinking are molded into higher-order cognitive functions.

A central concept of the theory is the *zone of proximal development*, which Vygotsky (1978) defined as 'the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers' (p. 86). Within this zone lies the greatest potential for learning because it contains abilities that are still forming and are on the cusp of emerging, but will not yet be manifested without social guidance:

> The zone of proximal development defines those functions that have not yet matured but are in the process of maturation, functions that will mature tomorrow but are currently in an embryonic state. These functions could be termed the 'buds' . . . of development rather than the 'fruits' of development.
>
> (1978, pp. 86–87)

Psychologist Jerome Bruner used the term 'scaffolding' to capture Vygotsky's ideas of how a child learns in social contexts, supported by more competent symbolic representatives of culture (Wood, Bruner, & Ross, 1976; see also guided participation, Rogoff, 1990). *Scaffolding* refers to support tailored to a child's current needs, constantly changing to fit the child's emerging abilities while learning a new skill. Thus, scaffolding requires close attention to what the child is capable of doing at every step of their learning and an awareness of the long-term goal of learning, which is to enable the child to function independently. By supporting skills that are currently forming or 'budding' within a child's zone of proximal development, the more skilled partner can provide a framework to stretch the child's abilities far beyond what he or she can reach independently.

Supporting musical exploration

What sort of scaffolding best supports young children's musical play and improvisation? Berger and Cooper (2003) observed children and parents for 10 weeks as they participated in a musical play program that included free play with a variety of musical instruments. The teachers' and parents' roles were 'not to instruct, but rather to be led by the child's exploration' (p. 154). With respect to adult behaviors that *enhance* musical play, the findings were in line with four techniques for scaffolding children's musical exploration, proposed by Fox (1989):

> (1) by simply watching a child making sounds, a parent lends value to the event through his or her attention; (2) by modeling a behavior . . . a parent becomes physically involved; (3) by making positive comments about the sound or action . . . a parent attaches verbal encouragement to the child's action; (4) by asking questions or making statements, a parent can offer vocabulary or ideas for constructing musical patterns. (pp. 15–16)

Indeed, under some conditions the mere presence of an interested adult may be sufficient to enhance play. When 95 children aged 2 to 4 years played with an Orff xylophone either in a solitary fashion or in the presence of an adult, the presence of an attentive and responsive adult sustained children's play for longer periods of time, and led to more inventive and complex explorations of the xylophone – even if the adult simply watched (Young, 2003b). Children spent the least time playing on a xylophone when alone (1 minute 40 seconds on average). They played almost twice as long when attended by an unfamiliar adult who was musically trained (3 minutes 12 seconds), and by far for the longest time when attended by a familiar adult with no musical training (6 minutes 49 seconds). It is interesting that the presence of musically untrained adults seemed to encourage the lengthiest play. Musically trained adults may have (unintentionally) constrained play by subtly encouraging and rewarding 'correct' ways of playing, thereby narrowing the range of exploratory behavior and diminishing the child's intrinsic curiosity. This finding is in contrast to Vygotsky's view that learning is facilitated by guidance from a more skilled partner. It is possible that at an early exploratory stage, the presence of a supportive, nondirective partner may most effectively scaffold learning.

The importance of child-led strategies in scaffolding children's musical play in collaborative contexts was further explored in another study by Young (2003c), who identified three styles of interaction that appeared to *stifle* children's interest in a turn-taking or matching game: (1) an adult taking a commandeering or directive role by telling children how to play instruments ('play it like this'); (2) an adult playing only his or her own melodies and rhythms, failing to match or incorporate the children's musical ideas; and (3) an adult engaging in poor turn-taking, such as starting to play too soon or too late to sustain or continue a group rhythm (p. 111).

In particular, these findings point to the importance of the adult imitating or elaborating on the *child*'s ideas. A child's creative efforts may be best supported when there is mutual effort towards coordinated playing as opposed to an approach in which only the child is expected to copy the adult. Examples of two types of imitation in collaborative musical play are 'sequential imitation' (one person plays; the other person listens and then imitates) and 'imitation to synchronize' (both play together while making an effort to match each other, by sending and reading each others' nonverbal cues). Both forms of synchrony are integrated into group improvisations among musicians, particularly in the jazz tradition.

The somewhat mixed findings about the effects of adult involvement in young children's musical play may be explained by the adult's style of involvement. As well-meaning adults often tend to be directive with children, many studies show minimal adult participation to be most effective in supporting musical play. However, a sensitive and unintrusive adult partner may facilitate more sustained, varied, and complex musical play, so that children are performing at higher levels within their zone of proximal development than when playing alone or unnoticed.

Beyond exploratory play: Orff, Dalcroze, and Kodály methods

As children advance in musical training and move toward specialized skills and knowledge in music, effective teaching requires more structure and direction. This guidance may be most effectively implemented if continuing to take a child-oriented approach, as exemplified by the *Orff-Schulwerk*, *Dalcroze*, and *Kodály* methods. All three methods are rooted in the assumption that children possess a natural aptitude for music. Basic musical concepts and skills are taught as a natural extension of ideas and behaviors that children have discovered and engaged in since infancy. The main themes of the present chapter resonate with these three methods, as all three approaches extend and refine the musical behaviors that young children spontaneously engage in during their play – singing, rhythmic speaking or chanting, movement, and improvisation.

In the Orff-Schulwerk method developed in the 1920s and 1930s by German composer Carl Orff (e.g., 'Carmina Burana'), for example, the voice is used as the first rhythmic instrument in 'speech play' or 'rhythmic speaking.' Orff viewed speech, chant, and song as points on a continuum, proposing that children move naturally from speech to various rhythmic activities and then to song. This is in line with Rutkowski's (1997) observations that young children's singing gradually progresses from a speech-like character to a singing tone. Words containing one, two, or three syllables are recited to develop a sense of pulse, meter, accent, and rhythm. Short repeated phrases are used to explore various simple forms, such as binary, ternary, and second rondo form. Next, the body is used to accompany the speech rhythm, through actions such as clapping, stamping, snapping fingers, and slapping hands on thighs ('patsching'). Finally, rhythms and melodies are created on instruments, beginning with unpitched percussive instruments such as claves, drums and tambourines, and later using pitched instruments such as Orff xylophones and glockenspiels that are specially designed for children.

Hungarian composer Zoltan Kodály also viewed the voice as the natural musical instrument, and encouraged children to learn a wide repertoire of folk songs from their own cultures. He opposed the teaching of simplified songs to children and believed that their earliest song models should be well-composed repertoire of high quality and lasting value. Kodály's ideas (disseminated in the early to mid-1900s) were developed into a teaching method by his students. Teachers following the Kodály method introduce pitches to children beginning with *so-mi* followed by *la*, consistent with some observations that the falling minor third and *so-mi-la* tones spontaneously emerge in the early songs and chants created by children. Rhythms are taught by starting with a steady quarter note (crotchet) and not the whole note (semibreve) as is often taught in traditional methods. The steady quarter note provides the rhythmic structure to which children march, while speaking rhythmic syllables (*ta* for the quarter note or crotchet, *ti-ti* for eighth notes or quavers, *teri-teri* for sixteenth notes or semiquavers). Thus, whole-body movement is used to feel the 'pulse' of music, bringing to mind Phillips-Silver and Trainor's (2005) study (from chapter 8) on the connection between 'beat' experienced as body movement and auditory perception of rhythmic structure. Hand gestures (Curwen hand signs) are used to show the relative position of pitches in space – thereby making scale relations visible through movement – while singing *solfège* (*do, re, mi*). As the syllables correspond to the degrees of a scale rather than the letter names of specific frequencies (e.g., C, D, E), the focus is on the relationship of pitches to each other and to *do* in any key.

'Listening through the body' is a central idea of the Dalcroze method, developed in the early 1900s by Swiss composer and music educator Émile Jaques-Dalcroze. The Dalcroze method employs 'eurhythmics,' in which the body is used as an instrument to express different elements of music such as dynamic range, timbre, pulse, rhythm, meter, phrasing, and tension-relaxation.

Whole-body movements as well as isolated movements are incorporated, especially to reflect sudden changes in musical dimensions. Consistent with Drake, Jones, and Baruch's (2000) and Gorali-Turel's (1999) previously mentioned finding that children's natural tempi and style of movements differ from those of adults, the initial lessons allow the children to move at their own tempi before being trained to adjust to an assigned tempo. Children gradually progress to synchronizing movements to peers, to matching sound and movement, and finally to improvising their own movements to music. Although the primary focus is on movement, the Dalcroze method also incorporates singing and playing with instruments.

Another theme of the present chapter that links to all three methods is their shared emphasis on the early introduction of musical improvisation. In traditional teaching approaches, the student is often taught to focus on reproduction and interpretation of musical works until reaching an advanced level of training. However, child-oriented methods target the inventive nature of the active learner. This strategy is consistent with the observational research reviewed in this chapter, suggesting that it may be easiest to support and reinforce these abilities before the child becomes more focused on the product (e.g., Brophy, 2005; Kratus, 1989). Further, the social context for these music methods is effective as young children often produce better compositions when collaborating with another child, especially a friend (MacDonald, Miell, & Mitchell, 2002).

The Orff-Schulwerk method introduces children to improvisation in a series of steps. First, children engage in variation of a familiar piece along one dimension such as dynamics, tempo, or rhythm, while keeping everything else constant. Improvisation and composition become increasingly novel and spontaneous, exploring many interacting dimensions of music. Those following the Kodály method focus on vocal improvisation, learning to notice patterns by listening and seeing Curwen hand signs in their learned repertoire of songs, developing variations (such as changing modes or rhythmic patterns of a familiar song), and finally creating their own improvisations.

The Dalcroze method begins with imitation of movement, but children eventually combine and invent their own expressive eurhythmic movements, rhythmic speech sequences, and rhythms and melodies on various instruments. Again, the idea of 'embodied rhythm' in both Kodály and especially Dalcroze methods resonates with Phillips-Silver and Trainor's (2005, 2007) research on the cross-modal interaction of movement and sound. Movement also expresses the form and structure of music in space, making physical and concrete what is usually abstract and fleeting as a musical piece unfolds in time. Advanced Dalcroze classes culminate in *plastique animée*, a technique whereby individual improvisations of students to express a musical piece (after multiple hearings) are discussed with a director, who scaffolds deeper analysis of the composition.

Skills such as learning to read and write musical notation are not typically introduced in the first lessons in any of these methods. Instead, children learn basic musical concepts by experiencing them first through movement, singing and playing instruments, and listening. Laying down the foundations of musicality before introducing notation may avoid producing what Schleuter (1997) refers to as 'button pushers' who simply view notation as a prescription for what fingers to put down without any real musical sensitivity for the sounds they are creating. This 'sound before symbol' approach draws on a long history of philosophies of music education (see, for example, Abeles, Hoffer, & Klotman [1994] on the views of Pestalozzi, Mason, and Mainwaring), and is in line with Bruner's (1966) idea that thought progresses from enactive to iconic to symbolic modes of representation.

In sum, many effective early childhood music methods incorporate techniques that expand on the ways young children spontaneously explore sound, and capture their natural inclination toward active, inventive engagement with sound. They assume the role of active learners who construct their own understanding of a world of sound through direct experience and guided discoveries in supportive learning contexts. The various methods are not intended to serve as an endpoint, but to lay the foundation for the development of more specialized vocal and instrumental skills through formal and informal means, and for a lifetime of engagement in musical activity.

Coda

The Orff, Dalcroze, and Kodály methods start with the assumption that all children have a natural aptitude for music, and that music education should serve to make music available to all – as opposed to the genetically 'lucky' or 'talented' few. Early childhood music education may be viewed as simply *continuing* the trajectory of naturally occurring behaviors that children already engage in, and reinforcing the proclivity to music that they already possess. Child-oriented music methods sustain and cultivate basic skills that are already in place, providing a bridge to more structured instrumental or vocal lessons that will continue the trajectory of musical development toward more specialized skills and technique. The next two chapters on music practice and performance will focus on the development and refinement of these specialized musical skills as one continues along the pathway toward musical expertise.

10 Practice and musical expertise

'How lucky you are to be able to play that instrument.'

'It's funny – the more I practice, the luckier I get!'

The second line of the exchange above is a popular quote that has been attributed to various professional golfers. In this context, it was quoted by Elizabeth Harré (former double bass player at the Sydney Opera House and daughter of author RH), responding to an admirer. Whether observing a virtuoso string concerto or a masterful stroke on the putting green, observers often casually attribute a great performance to luck or inborn talent. But as anyone who excels in any field knows, the gap between innate capacities and mastery is filled with years of hard work and training!

While the last two chapters have focused on the emergence and development of basic musical abilities, we now shift our attention to the refinement of more specialized skills. Up to this point, we have shown how infants and children appear to have a natural propensity for music. However, specialization in the field requires much more than this. Studies show that in the early years, informal experience and exposure to music is sufficient to develop some basic musical capacities. For instance, Lamont (1998) found that from about 6 to 11 years, children performed quite similarly on 'probe-tone' tests (in which the listener must evaluate how well a note follows the tones before it [chapter 5]), regardless of whether or not they were taking music lessons. However, pitch perception did not improve significantly after age 11 years without more intensive involvement in music, usually in the form of instrument or vocal lessons.

An essential element of this active involvement in music is the hours spent in individual or group music practice. The present chapter focuses on one of the most practical topics relevant to a musician's life: practice. Until quite recently, there was little research on music practice. Studies on music performance tended to focus on the product, rather than the process, of learning a musical work. It is only since the 1980s or so that a body of literature has emerged on this topic (Gruson, 1988). In this chapter, we review classic and current studies on music practice. The core of the discussion explores effective practice strategies, but we shall begin with a discussion of two forms of music practice: deliberate and informal.

Deliberate and informal practice: The acquisition of expertise

Deliberate practice

The phrase *deliberate practice* refers to highly structured practice with the explicit goal of improving performance, and involves tasks designed to address weaknesses and correct errors (Lehmann & Ericsson, 1997). In the domain of music, deliberate practice is exemplified by the hard work involved in improving the performance of scales, études, and musical pieces assigned by the teacher by using techniques to correct mistakes and strengthen performance. It is not the same as *mere repetition*, as deliberate practice involves specific strategies to bring about change and reorganization in behavior, as opposed to simply repeating the same actions with the hope that it will automatically lead to improvement.

While mere repetition may be sufficient to bring about improvement in the early stages of learning a skill, progress is likely to be arrested before reaching high levels of proficiency if relying mainly on this method. In some respects, practicing without concentration may even interfere with progress. McPherson and Renwick (2001) found that children spend over 90 percent of their practice time simply playing through their pieces from beginning to end, without intentionally setting specific goals or engaging in the self-monitoring required to correct mistakes. Even older students often do not correct mistakes when practicing. In a study by Hallam (1995), approximately 60 percent of 6- to 18-year-old beginning to advanced string players left some errors uncorrected in their solo practice. However, unless conscious effort is made to correct errors and improve performance as is characteristic of effective deliberate practice, errors may be 'practiced in' through repetition (Gruson, 1988).

In a widely cited paper published in 1993, K. Anders Ericsson and colleagues argued that the role of innate capacities in achieving superior performance has been overemphasized, and that the effects of 'deliberate practice' may be much more significant than previously believed (Ericsson et al., 1993). Arguing against the notion that level of achievement is determined primarily by genetic endowment (e.g., Galton, 1869/1979), their central claim was that 'the level of performance an individual attains is *directly related to the amount of deliberate practice*' (p. 370, emphasis added). Ericsson and his colleagues devoted many years to examining the role of deliberate practice in the acquisition of expertise in a wide range of fields, such as sports, music, arts, and sciences. In their thorough review of the literature, they found little evidence for 'heritable characteristics' that served as reliable predictors for superior performance in a field. Instead, they found accumulating evidence for a direct link between deliberate practice and exceptional levels of performance.

In the domain of music, Ericsson et al. (1993) studied three groups of violin students enrolled at a highly selective music conservatory in West Berlin: 'the best violinists,' a group of 'good violinists,' and a third group who met a lower standard of achievement. Estimates of hours of solo practice were computed from biographical data (see example in Table 10.1, which may inspire you to review your own practice history and accumulated hours of practice). Consistent with their claim, the researchers found that the 'best' violinists had accumulated an average of about 7400 hours of solo practice by the age of 18 years, compared to 5300 for the 'good' violinists and 3400 for the lowest-achieving group. Ericsson and colleagues also examined pianists at the same prestigious music conservatory, and found that 'expert' pianists had accumulated an average of approximately 7600 hours of practice by age 18 years, compared to only about 1600 by amateurs.

Similar findings have been yielded in other studies comparing hours of accumulated practice in groups of varying levels of music achievement (e.g., Sloboda, Davidson, Howe, & Moore, 1996; though see Madsen, 2004 and

Table 10.1 Retrospective estimate of accumulated hours of deliberate practice by an expert pianist (Krampe & Ericsson, 1995, p. 90)

Year	Age	Event	Estimated practice
1941	5	Start practice, school	1.5 hours × 7
1942	6		
1943	7		
1944	8		3 hours × 7
1945	9	First major public concert	4 hours × 7
1946	10		
1947	11		
1948	12		4 to 6 hours × 7
1949	13		
1950	14		
1951	15	End of instruction from father	
1952	16	Finish school, music college preparation	
1953	17	Start music college	6 to 8 hours × 7
1954	18		
1955	19	Participation in Chopin competition	
1956	20		
1957	21	Graduation from music college	
1958	22		
1959	23		
1960	24		
1961	25	Birth of first child	
1962	26		6 hours × 7
1963	27		

Source: Reproduced with permission from the authors and Cambridge University Press.

Williamon & Valentine, 2000 as the relationship between amount of practice and achievement may be weaker at the sub-expert level). Such findings challenge the popular belief that genetically 'gifted' or 'talented' individuals naturally progress to higher levels of performance more quickly than most. According to Ericsson and colleagues, the difference between the groups of different levels of musical achievement is primarily due to the greater amount of practice accumulated over the same period of time. Of course, one must also take into account the question of whether personal variables such as motivation and temperament, which may be influenced by one's genetic make-up, may account for why some people practice longer and more effectively than others.

Short-term engagement in deliberate practice, no matter how effective and intense, may not be sufficient to acquire expertise. Simon and Chase (1973) first discovered that very few players had reached the level of international chess master without at least 10 years of intense involvement in chess. Since Simon and Chase's classic paper, the so-called *'10-year rule'* has been supported by numerous studies examining eminent achievement in a variety of disciplines including mathematics, science, poetry, literature, tennis, swimming, long-distance running, and interpretation of X-rays (see Ericsson et al., 1993, p. 366, for a review). Some studies suggest that more than 10 years of intense deliberate practice are required to achieve eminence in certain fields – including music. For instance, Sosniak (1985) found that concert pianists had studied piano for an average of 17 years before qualifying for the finals of their first major international piano competition, and Hayes (1981) and Simonton (1991) found that music composers had undertaken about 20 years of music study before producing the first eminent composition.

That is not to say that intensive training alone accounts for exceptional levels of achievement. The deliberate practice model involves a confluence of variables that interact in complex ways to produce individual outcomes. Social context plays a significant role as exceptional achievement is rarely attained in isolation. For instance, Gary McPherson (2009) has proposed a feedback loop model of the role of parents on children's musical development. To summarize the model: The particular goals that parents have for a child's music education (which is shaped by their beliefs, values, and ideals) influence the parenting styles and practices used in their daily interactions with the child concerning their musical development. However, the influence does not only flow in one direction, but is bidirectional in that the characteristics of the child and larger sociocultural variables surrounding the child mediate the effects of the parents' influence – and in turn also shape the parents' behaviors toward them – thus creating a feedback loop that may lead to many possible outcomes. One might assume a similar feedback loop between teachers and pupils.

With respect to the child's characteristics, level of intrinsic motivation (as opposed to engagement for extrinsic purposes such as rewards) and sense of *self-efficacy* or a person's belief about their capacity to perform a task

(Bandura, 1997) are central to musical achievement. Another important factor emerging from recent research in education is the role of metacognitive strategies (which can be broadly defined as one's ability to monitor the cognitive processes involved in learning). For instance McPherson's (2005) longitudinal study followed 157 children (aged 7 to 9 years) from the time they began instrumental lessons (mainly clarinet, trumpet, flute, or saxophone) through the next 3 years. One of the main findings was that accumulated practice hours explained between 9 and 32 percent of the variance in a measure assessing the ability to play rehearsed pieces from notation, but the quality of *metacognitive or mental strategies* was a more powerful predictor for the children's performance on sight-reading, playing from memory, and playing by ear. The higher achieving students tended to use more sophisticated mental strategies in practicing, sight-reading, memorizing, and playing by ear. For instance, with respect to strategies for practicing, these students were more likely to keep a practice diary to keep track of what they learned, to self-correct, and to apply specific techniques to improve their skills. While many of the children simply went about practicing out of habit, McPherson concluded that the high achievers in the study engaged in 'deliberate practice' as characterized by Ericsson and colleagues.

The contribution of genetics is also not ignored in the deliberate practice model (e.g., see Lehmann & Ericsson, 1997). As other researchers have pointed out, there is some indirect evidence for atypical brain organization in gifted children, and the effects of deliberate training alone may not be sufficient to wholly account for exceptional abilities identified in very young children with only a few years of experience in their domain (e.g., see Winner, 2000). Further, some of the variables discussed earlier – such as a child's temperament and level of persistence – may also stem from the genetic make-up of the child. Nonetheless, it appears that the role of training and effort plays a far greater role in the acquisition of expertise than previously assumed.

Informal practice

In contrast to deliberate practice, *informal practice* refers to practice time not spent on assigned exercises and pieces, but on 'playing for fun' (Sloboda et al., 1996). It includes playing favorite pieces that are no longer assigned, sight-reading music for enjoyment, playing by ear, improvising and other creative musical activities, and just 'playing around' and experimenting with sounds and motor patterns. Although informal practice can involve self-correction and improvement over time, the goal is often more immediate (to enjoy the activity). In the previous chapter, we explored the ways in which play provides an important context for learning skills that cannot be readily acquired in more structured contexts. The case of informal practice suggests that learning through play may not be limited to the childhood years. Further, playful involvement in music brought about by informal practice may sustain interest in music study, as there is evidence that students who engage in informal practice continue lessons longer than those who do not (McPherson & McCormick, 1999).

Unfortunately, informal practice has often been neglected in discussions on music practice and has received little attention in the literature (Sloboda et al., 1996). Students are often discouraged from engaging in informal practice and told to 'stop messing around and get back to your lesson.' However, informal practice builds essential musical skills. In particular, it develops expressivity, as one is more likely to play 'with feeling' when less focused on the technical aspects of performance (Sloboda, 2005, p. 270). Activities such as playing by ear and improvising develop auditory imagery (i.e., the ability to anticipate the sound of tones before they are played) which in turn has been linked to good sight-reading ability, especially as repertoire increases in complexity (Kopiez & Lee, 2006). Informal practice with others, such as in jam sessions, builds flexibility and coordination as one must respond quickly to musical ideas as they are introduced, and learn to synchronize and harmonize with others.

People often seem to assume that playing by ear, arranging, improvising, and composing are skills that are naturally 'picked up' by the truly musical – and overlook that they too require practice! Jazz pianist Glenn Tucker (shown in Figure 10.1) spends about two thirds of his practice time on technical

Figure 10.1 Emerging jazz pianist Glenn Tucker spends two thirds of his practice time on technical exercises and repertoire and one third on informal practice, mainly improvising at the piano. Tucker studies with Geri Allen and has performed with James Dapogny, Andrew Bishop, and John Clayton.

Source: Photograph used with permission of Glenn Tucker and PKO Records. Copyright © John Lilley.

exercises, transcriptions, and classical repertoire, and one third of the time on informal practice, mainly improvising at the piano: ‘I think that one of the biggest goals in jazz is to be spontaneous, and the only way to be spontaneous is to practice being spontaneous’ (personal communication). Explaining his approach to informal practice, he says: ‘I set parameters for myself, like a key or pedal point, a metronome marking, or even a style or era – 1920s Harlem stride, classical sonata form, bebop, or sometimes the parameter is to improvise without meter or tonality. However, these are more like crutches; the real idea is to start with silence and do a stream of consciousness free improvisation, and to go where the music takes you.’ In general, it appears that formal and informal practice develop different aspects of musicality, and involvement in both types of practice may produce the most well-rounded musicians.

Striking the balance

In one of the few studies examining both deliberate and informal practice, Sloboda and colleagues (1996) examined the practice habits of five groups of music students with differing levels of achievement. Hours of practice were estimated by studying diaries and conducting interviews with over 250 students and their parents. The researchers found that by age 12 years, the most advanced children had accumulated over 800 percent as much practice time as the least advanced students (as discussed in Sloboda, 2005). By age 17 the estimated total practice time for the best pianists was just under 7000 hours, which is in step with Ericsson et al.’s (1993) estimates for the ‘best’ violinists and ‘best’ pianists. In addition, Sloboda et al. (1996) found that the top two groups also devoted more practice time on average to *informal* practice than did the two lowest-achieving groups (although estimated hours of informal practice were not reported in the study).

Because so few studies have addressed informal practice, it is not yet known what the ideal ratio of the deliberate to informal practice might be. However, as deliberate practice is most effective for improving technique and precision, it is common wisdom that students should spend significantly more time on deliberate than informal practice, particularly on exercises. In Sloboda et al.’s (1996) study, the top achieving group not only spent far more time practicing, but also dedicated a larger proportion of practice time to scales and technical exercises than did the other four groups. Of course, the ideal proportion of deliberate and informal practice may also vary with genre, for instance for jazz musicians who must draw more heavily on improvisational skills.

Practice strategies: A review of the research

Many students are not very deliberate or strategic in their practice. As one honest school-aged participant told researcher McPherson: ‘If I’m going really well I play it a few times. If it's bad then I only play it once, except I know it should be the other way around' (2005, p. 18). In this section, we will discuss a few themes in the classic and current literature on music practice strategies: distributed versus massed practice, part versus whole approaches, mental versus physical practice, and aural models versus verbal instruction. The focus will be on *deliberate practice*, although some principles may also apply to informal practice. Most of the research in this area has concentrated on solo practice at the keyboard. Although they will also be reviewed, there are far fewer studies on music practice involving other instruments, voice, and ensembles.

Distributed versus massed practice

The issue of 'distributed versus massed practice' refers to whether practice time is more effectively divided into several short sessions (distributed), or concentrated into a single session (massed). Several studies show distributed practice to be more effective than massed practice for learning many types of motor skills (Oxendine, 1984). In the area of music learning, studies also show benefits of distributing practice into short sessions throughout the day. In a series of early studies, for instance, Rubin-Rabson (1940a) found that massed practice leads to quicker – but not necessarily deeper – learning of short musical sequences. Students were able to learn music faster through intense sessions of massed practice. However, distributed practice led to better retention of the music after two weeks, and more accuracy when asked to transcribe the music from memory. Rubin-Rabson speculated that breaks between practice sessions give students an opportunity to take a fresh look at the music, and to understand its structure more deeply.

Subsequent studies have also shown that many concert performers and advanced music students use distributed practice. In Ericsson et al.'s (1993) study described earlier, even the 'best' violin students enrolled in a top conservatory engaged in an average of only about 3.5 hours of solo practice per day (not including ensemble practice). The 'best' violinists distributed their practice into three to four sessions per day that varied from approximately half an hour to about 2 hours per session. They usually took two short breaks within each practice session, often playing for no more than about 50 minutes before taking the first break (reported in Krampe & Ericsson, 1995). It is interesting to note that the top two groups of violinists also napped more often than the less accomplished third group (Ericsson et al., 1993, p. 375). This finding is in line with research on the critical role of sleep – independent of whether it occurs at daytime or nighttime – in *consolidating* memory for motor skills (Fischer, Hallschmid, Elsner, & Born, 2002). It appears that the memory trace for motor learning continues to be reprocessed even during periods without intervening training, and that sleep plays a critical role in this 'slow component of learning' (Maquet et al., 2003).

Part versus whole approach

'Part versus whole' pertains to the question of whether musical works should be divided into smaller segments during practicing, or practiced as an intact whole. One limitation of early studies was that researchers often gave participants instructions on how to segment the pieces. For instance, in Rubin-Rabson's (1940b) study, students were told to learn a composition as a whole, divide the composition into four-measure phrases and learn each phrase before moving on, or divide phrases into smaller parts and learn each part before moving on. While giving researchers a high level of control, this procedure did not reveal what 'part' or 'whole' strategies students use when left to their own accord.

Subsequent studies have examined ways in which participants practice without specific directions. In Mishra's (2002) study, for instance, students were simply told to learn a 36-measure étude. The following strategies spontaneously emerged:

> *Serial:* Students using this approach tended to return to the beginning of the piece every time a mistake was made. There were usually many returns to the starting point, and the last part often underwent little practice.
>
> *Segmented:* This approach usually began by dividing a composition into segments, and learning each segment separately. The separate segments were then linked together and the joints between segments were carefully practiced.
>
> *Additive:* Like the segmented approach, the work was divided into sections, but continually lengthened so that the piece was learned in rapidly expanding segments. There was usually a deliberate effort to play the whole piece as soon as possible.
>
> *Holistic:* This approach entailed playing pieces from beginning to end, trying to avoid stopping and restarting. If a mistake was made, the student would backtrack just enough to correct the error, but not to the very beginning.

Although all four approaches resulted in the successful memorization, some were more efficient than others. Out of the 80 instrumentalists in Mishra's (2002) study, the four who took the longest time (66 to 100 minutes) to memorize the 36-measure étude used Segmented or Serial approaches. The instrumentalists who took the shortest time (8 to 16 minutes) used Holistic or Additive approaches. Clearly, the most efficient memorizers used methods that focused on 'whole' rather than 'part' learning.

However, it is possible that Holistic and Additive approaches may be superior only when learning and memorizing relatively short and simple musical works. Case studies of professional musicians preparing much longer and more complex works have shown that they often rely on a combination of 'part' and 'whole' approaches in their practice routines (e.g., Chaffin & Imreh, 2001; Hallam, 1995; Miklaszewski, 1989). Experts often divide extensive musical works into structural units to be practiced in Segmented or Additive fashion. But within those sections, they often use a Holistic approach, repeating parts that need improvement but pushing forward quickly through the section.

In one of the most thorough analyses of music practice in experts, Chaffin, Imreh, and Crawford (2002) examined how concert pianist Gabriela Imreh prepared a movement from Bach's *Italian Concerto* for performance, by analyzing videotapes of her practice sessions (see also Chaffin, Lisboa, Logan, & Begosh [2009] for an extensive analysis of a cellist's practice). The authors devised elaborate visual representations to convey Imreh's approach to practicing the piece, as shown in Figure 10.2. The figure represents approximately 15 minutes of the pianist's first practice session on the portion of the musical work shown on the stave. The gray dotted lines depict what Chaffin and colleagues refer to as 'runs' (playing through longer segments to see how the parts fit together) and the solid black lines represent 'work' (repetitions of

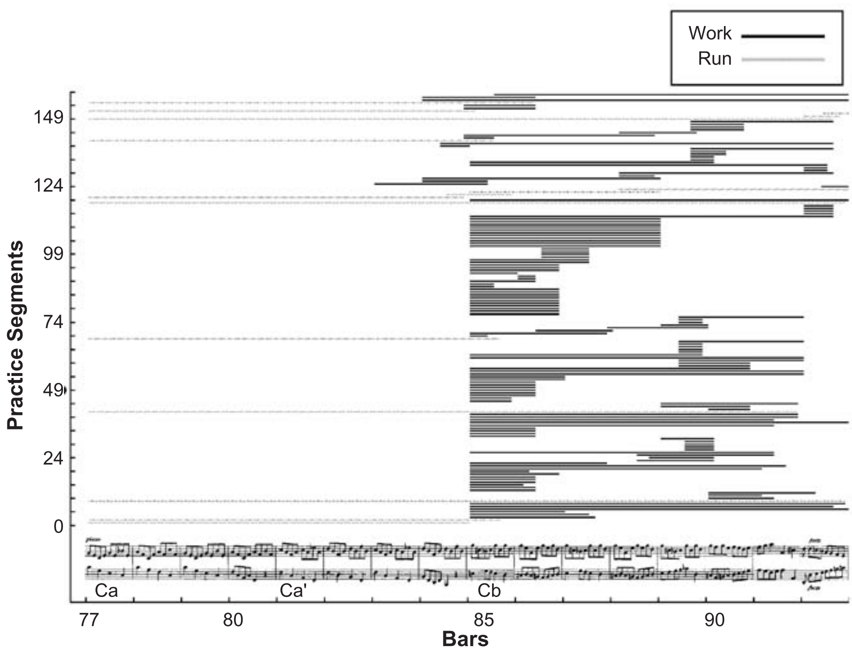


Figure 10.2 Practice record for the first practice session of a section of the *Italian Concerto (Presto)* by J. S. Bach (Chaffin & Imreh, 2001, figure 2, p. 47).

Source: Reprinted by permission of Sage.

short passages). The figure should be read from left to right, and from the bottom to the top. As shown, Imreh began by 'running' through sections Ca to Cb, and then 'worked' on a few measures repeatedly at the beginning of Cb before returning to Ca to play through a long stretch of the music. There were many returns to work on small parts of Cb, interspersed with longer runs. One can see the pianist's use of both 'part' and 'whole' strategies. Chaffin et al.'s observations of Imreh's practice also typify good deliberate practice: 'everything was done with a specific goal in mind' (p. 118).

Chaffin and colleagues' (2002) study of Gabriela Imreh's practice techniques also showed that she segmented the music into parts corresponding to the musical structure of the work. Advanced students and experts tend to be more deliberate about how they segment the music, dividing the musical work into meaningful structural units such as themes, phrases, sections, and bridge passages (Miklaszewski, 1989; Williamon & Valentine, 2000). They also tend to learn the units in a structurally meaningful order, such as principal themes first and then variations, not necessarily approaching the parts in chronological order. By contrast, younger and beginning students often simply segment the composition at points of convenience – for example, learning the music line by line (e.g., Hallam, 1997).

When it comes to learning vocal repertoire, a different sort of 'part versus whole' question might arise. Is it more effective to learn text and music separately first, or together from the start? In previous studies, participants often reported that they learned music and words separately, but these studies relied on data gathered from interviews. Ginsborg (2002) undertook an observational study of 13 classically trained singers in order to find out how singers actually practice. The singers were observed over the course of four to six practice sessions of 15 minutes each, as they learned an unfamiliar song. The findings showed that fast, accurate memorizers spent more time practicing words and music *together* during the early (first two) sessions of their practice than slow, inaccurate memorizers. Also, fast, accurate memorizers spent more time practicing music and words *separately* during the final two to four practice sessions than did the slow, inaccurate memorizers. In sum, the most effective technique was to learn words and music together initially, and then study the parts separately. A subsequent study showed that memorizing words and melody together is an effective strategy for accurate and fluent recall of songs, but only for more advanced singers (Ginsborg & Sloboda, 2007).

Mental practice techniques

Mental practice refers to 'the imaginary rehearsal of a physical skill in the absence of gross [i.e., large] muscular movements' (Theiler & Lippman, 1995) and is widely used by athletes. In music training, it refers to practice that takes place in a person's mind while imagining the sounds and motor actions involved in playing a musical work. Two main forms of mental practice techniques have been described: analytical pre-study and mental rehearsal.

Analytical pre-study consists of studying the musical score *before* engaging in physical practice (Rubin-Rabson, 1941a). This allows students to preview the musical work from beginning to end – analyzing its form, identifying patterns, and 'hearing' the musical work in their heads – before physically playing the piece. Both analytical skills and auditory imagery are involved, as the performer analyzes the composition's structure and imagines how the notes will sound. The performer also visualizes the particular movements required to play or sing the music, therefore identifying technically demanding parts and imaging specific motor actions (e.g., working out fingering, imagining position of bow or change of embouchure). As we shall discuss in the last section of this chapter, several studies show responses in the auditory cortex in the absence of physical sound and action, when a person is only imagining the sound of music or imagining the actions used to produce it (see Zatorre & Halpern, 2005).

Engaging in analytical pre-study prior to physical practice may facilitate smoother, more accurate, and less mechanical playing when sight-reading or playing through a piece for the first time. Good sight-readers usually engage in pre-study before playing. In particular, pre-study can facilitate rhythmic accuracy when playing complex pieces because the student has time to work out the rhythms first (Rosenthal, Wilson, Evans, & Greenwalt, 1988). Any technique that can improve rhythmic accuracy facilitates performance, as the most common errors in sight-reading are related to rhythm (McPherson, 1994).

In the book *Piano Technique: The Shortest Way to Pianistic Perfection* (Gieseking & Leimer, 1932/1972), piano teacher Karl Leimer and his most celebrated student pianist Walter Gieseking endorsed a more intensive form of analytical pre-study. Using the Leimer-Gieseking method, a student focuses on pre-study of a composition for days or weeks without playing the piece (though engaging in physical practice of other assigned pieces). After memorizing the work through careful analysis, the student then performs extended passages of the music immediately from memory – or, in the case of Gieseking, sometimes whole musical works!

Others recommend that the pre-study period should be intensive, but brief. Rubin-Rabson found that only 5 or 6 minutes of pre-study is sufficient to facilitate memorization of keyboard music, and retention after a 2-week period (Rubin-Rabson, 1941a, 1941b). Extending pre-study time to 9 or more minutes did not seem to significantly improve performance, perhaps because attention and concentration begin to wane.

Another form of mental practice is *mental rehearsal*, implemented *after* engaging in a period of physical practice, once the performer already knows how the piece 'feels under the fingers.' (Although it is also referred to as 'mental practice,' we will use Rubin-Rabson's [1941b] term 'mental rehearsal' to distinguish it from analytical pre-study.) Mental rehearsal involves playing through parts or all of a musical work in one's mind (with or without the score), using one's direct experience with the piece to imagine each movement and sound as vividly as possible. It may also involve visualizing the musical score, along with the sounds and motor actions of playing. The main aims of mental rehearsal are to increase the level of mental concentration, refine the motor actions, and solidify memory by filling in 'gaps' in one's memory of the musical score.

In general, physical practice appears to be a more effective means of improving performance than mental rehearsal alone, especially at the early stages of acquiring musical skills (e.g., Lim & Lippman, 1991; but see Cahn, 2008). However, some studies have found a combination of physical practice and mental rehearsal to be equally beneficial in improving performance for both instrumentalists and vocalists as physical practice alone (Coffman, 1990; Ross, 1985; Theiler & Lippman, 1995). For instance, Ross showed that improvement in pitch, rhythm, and articulation was similar for slide trombone players using a combination of physical and mental practice, or physical practice alone. This is good news for performers who must rehearse when musical instruments are not available, or performers who are injured and need physical rest.

A few studies have found that under some conditions, mental rehearsal interpersed with physical practice may even be more effective than physical practice alone. Specifically, combined mental and physical practice may be superior to pure physical practice for developing tonal quality in singers and pitch accuracy in guitar players (Theiler & Lippman, 1995), and aiding memorization of keyboard music in advanced students (Rubin-Rabson, 1941b). Further, a technique called 'shadowing' (making motor actions without producing sound – such as moving fingers lightly on the surface of piano keys, or moving the slide soundlessly on a trombone) may also be somewhat more effective than pure mental rehearsal (Ross, 1985). Interestingly, this technique may be more effective for performers who are better able to engage in auditory imagery (Highben & Palmer, 2004).

Although mental practice requires time and effort, it may speed up learning and memorization (Rubin-Rabson, 1941a). Even a single session of 4 minutes of analytical pre-study appears to be sufficient to facilitate memorization (Rubin-Rabson, 1941b), and mentally playing through a musical work just one time can already improve performance (Ross, 1985). However, mental practice may not be as effective for beginners or young children, or for repertoire that is large in scale or complex. It should also be noted that the studies reviewed above employed only short-term exposure to mental practice, and observed only its short-term effects. Mental practice itself takes practice, and future studies are needed to examine the possible long-term effects of pre-study and mental rehearsal.

Modeling

Modeling is a teaching tool in which a specific behavior is imitated by a student. A choir director may demonstrate a new vocal technique for the choir to emulate, or a conductor may clap a rhythmic passage for the ensemble to clap or play back. Until the invention of recording technology, music students were dependent on live models. Today, teachers can equip students with aural models (audio or audiovisual recordings) to use during solo or group practice. Jazz musicians deeply incorporate aural models into their training. Glenn Tucker (shown earlier in Figure 10.1) states, ‘I think that repeated focused listening might be the most thorough way to learn about music. On more than one occasion, I have found myself in an improvisational moment that calls for something I’ve heard on a recording, and more often than not I can directly reference a moment in recorded jazz without having physically played it before’ (personal communication). As an example of focused listening, he points to John Coltrane’s recommendation to listen repeatedly to a piece, each time attending to a different instrumentalist.

Aural models appear to be very effective for both solo and group practice, and students of different levels of training. In two studies employing advanced instrumentalists, listening to a recording of a musical work without doing any physical practice led to performances that were roughly equivalent (Rosenthal et al., 1988) or even superior to (Rosenthal, 1984) pure physical practice on measures such as correct notes and rhythms, accuracy of phrasing and dynamics, and precise articulation. Less advanced high school band members also showed greater improvement in rhythmic accuracy and playing at the assigned tempo when practicing with an aural model than without a model (Henley, 2001).

Further, the model does not have to be instrument-specific to be effective. Rosenthal and Henley’s studies both used an audio recording of a violin performance as the aural model for brass and woodwind instrumentalists. Most students seem to be able to transfer the skills to their respective instruments (Dickey, 1991), perhaps with the exception of beginning students (Linklater, 1997) and young children (Green, 1990). Young children, in particular, may need to follow a model that is as similar as possible to their own voices or instruments. For instance, Green found that first- through sixth-grade elementary school children were most accurate in imitating sung pitches when following a child vocal model, less accurate when imitating a female adult vocal model, and performed least accurately when the vocal model was an adult male.

Audiovisual models (such as videos and films) can also be effectively incorporated into solo and group practice sessions. For example, Linklater (1997) exposed beginning clarinet players to one of three modeling conditions during 8 weeks of practice: a videotape of a clarinet performance, an audiotape of the same clarinet performance, or an audiotape of other instruments playing an accompaniment (as a control). After 8 weeks, students in the clarinet-videotape model group scored significantly higher than the other groups when judged on several aspects of technique (embouchure, hand position, instrument position, and posture) immediately after treatment, and also scored higher on tone quality and intonation when evaluated a few months later.

Exact reproduction of the model may be the aim for beginning students. However, more advanced students should become progressively more independent in making their own decisions regarding what features to imitate, and to construct their own interpretation by listening to many different recordings. In this way, independent musicianship is encouraged, and students are less likely to become dependent on models (Brooks, 1995). In fact, Rosenthal (1984) found that advanced instrumentalists show greater improvement when provided with an aural model alone, compared to an aural model accompanied by detailed verbal instructions about where to allocate attention. She concluded that advanced instrumentalists already have a good sense of what they need to focus on when listening to a model to prepare a particular piece, and may be hindered by extraneous verbal guidance.

Music practice and the brain

Music practice is of interest to neuroscientists because of its implications for *neural plasticity* (discussed in chapter 4), the impressive capacity of our nervous system for change. The idea that intensive motor training can induce *functional* changes in the brain (i.e., modification of the organization of neuronal networks to increase efficiency) has been quite well documented (e.g., Pantev, Engelien, Candia, & Elbert, 2003). Recent studies have found evidence that increased use can also lead to *structural* changes in the brain (Pascual-Leone, 2003). In other words, training may not just bring about change on the level of neuronal reorganization, but may also guide the way the brain develops. In the final section of this chapter, we consider whether musical practice brings about structural and functional change in the brain independently of one's genetic make-up.

Effects of practice on motor control

When considering plasticity it is important to address whether musicians are born with atypical brains that predispose them to musical ability, or whether intense training over a long period of time can induce structural adaptations in the brain. Although research in this area is relatively new, there is accumulating evidence that structural changes are due to 'use-dependent' changes rather than differences present at birth. For instance, Gaser and Schlaug (2003) found a significant correlation between level of music training (namely, professional musician, amateur musician, and nonmusician) and amount of increase in gray matter volume in the motor (precentral gyrus), auditory (Heschl's gyrus), and visual-spatial (parietal) regions. Specifically, professional pianists had the highest gray matter volume in these regions, and nonmusicians had the lowest volume. The researchers attributed their finding of higher gray matter volume in professional musicians to 'structural adaptations in response to long-term skill acquisition and the repetitive rehearsal of those skills' (p. 9240), rather than to innate differences.

Schlaug and colleagues have also undertaken a longitudinal study focusing on a group of 5- to 7-year-old children, starting *before* they began music training (Schlaug, Norton, Overy, & Winner, 2005). Preliminary findings showed 'no pre-existing cognitive, music, motor, or structural brain differences' before instrumental training began for children who subsequently did or did not undertake instrumental training (p. 223). However, after a year of training, children who chose to take piano or violin lessons performed significantly better on tests of fine motor skills and auditory discrimination skills when tested on Gordon's Primary Measures of Music Audiation than a control group of children who did not undertake training. Moreover, the authors observed possible evidence for structural changes in the brains of children who took music lessons (specifically, nonsignificantly greater increase in gray matter volume), as compared to those who did not. To date these findings are preliminary as not all participants had completed their tests at the time of the report.

The findings of a few studies suggest that some structural adaptations may be instrument-specific. For example, Elbert and colleagues found that the region of the somatosensory cortex that represents input from the left hand was significantly more responsive to tactile stimulation in a group of expert string instrumentalists than in nonmusicians (Elbert et al., 1995). This may be because the fingers of the left hand may be more dexterous among professional string players due to training. Further, the effect was more pronounced for the string instrumentalists' left pinkie finger than for the left thumb. This may be due to the fact that the thumb is used mainly to hold the neck of the instrument whereas the fingers (including the pinkie) are constantly in motion (as shown in Figure 10.3). Elbert and colleagues attributed the increased responsiveness of the activated area to the possibility that the brain assigns more neural tissue to the processing of the hand and fingers that are more heavily used. In contrast to players of stringed instruments, pianists may favor the right hand. Bangert and Schlaug (2006) investigated the shape of instrumentalists' precentral gyrus, the fold that separates the motor cortex from the parietal lobe (see Figure 4.2). The shape of this fold suggested greater density of the motor area in the left hemisphere (which controls the right hand) for pianists, and greater density of the motor area in the right hemisphere for violinists.

Further, the most pronounced differences have been found for musicians who commenced training at a young age, most likely because intense training took place during a time when the brain was particularly malleable. In a study by Schlaug and colleagues (Schlaug, Jäncke, Huang, Staiger, & Steinmetz, 1995) the size of the anterior corpus callosum was larger in a group of professional musicians than in a group of matched controls, but this difference was larger for a subgroup of their musicians who had begun music training at 6 years or younger. The development of this anatomical change has been confirmed in subsequent research that involved measuring structural changes in a group of children before and after 15 months of musical training

Figure 10.3 Neuroscientists have found evidence of increased responsiveness in the region of the somatosensory cortex that represents input for the left hand (especially the pinkie finger) in expert violinists, after years of playing an instrument which requires dextrous movements of the fingers of the left hand and particularly the little finger.

Source: Photograph by John Lacko, used with permission of Barry Ross and the Kalamazoo Symphony Orchestra. Copyright © Kalamazoo Symphony Orchestra.

(Schlaug, 2009). However, although effects seem to be greater for musicians who started music training early in life, differences are sometimes found for 'late starters.' In Elbert et al.'s (1995) study, string players who had not started music study until after age 12 still showed higher levels of electrical activity in areas corresponding to the left pinkie finger, compared to non-musicians. As with so many areas of human development, it appears that there is a more flexible 'sensitive' (rather than rigid 'critical') period for the brain to adapt to intense motor training.

Effects of practice on music perception

Practice may not merely affect the motor areas involved in performing music. Recent evidence suggests that practice may also influence neural processes involved in the perception of music. Performers' brains respond differently to music than nonmusicians', possibly because listening to music that one knows how to perform elicits 'mental simulation' of the performance itself.

Mental simulation refers to imagining the production of an action without physically performing it, and is the foundation of mental practice (discussed earlier). Research in this area has recently surged. In the 1990s, researchers discovered (by accident) that monkeys' brains possess cells that have since been dubbed *mirror neurons* (Rizzolatti & Craighero, 2004). Specifically, when doing research on neurons that regulate motor behavior (in this case, reaching for a food pellet), Rizzolatti and colleagues found certain cells that responded similarly regardless of whether the monkey reached for the pellet or observed an experimenter reaching for the pellet. Since that time researchers have speculated that the role of these cells is for mental simulation. Subsequent research suggests that such neurons exist in humans and apply to hearing as well as performing music.

Although mental simulation of music in the brain has been documented for musicians and nonmusicians (Zatorre & Halpern, 2005), musical training appears to enhance the degree to which hearing music elicits simulation of production. Recent research on this topic is remarkable in that changes in neural responses have been found after only 20 minutes of practice among individuals with no prior musical training. In one study, participants were presented with melodies they had just practiced as well as melodies that resulted from simply re-arranging the order of pitches in previously learned melodies (Lahav, Saltzman, & Schlaug, 2007). Neural responses were greater, specifically in Broca's area, when listening to previously learned melodies that the participants had practiced. (It should be noted that Broca's area was once thought to be a 'language area,' though it is now considered to serve a more general role as part of the mirror neuron system mentioned above.) In another study, only 20 to 30 minutes of practice elicited larger neural responses (in similar areas) to melodies that had been practiced versus melodies that had been learned merely through listening without playing (Mutschler, Schultze-Bonhage, Glauche, Demandt, Speck, & Ball, 2007). Performers thus build associations between actions (key presses) and their effects (pitches) rapidly.

Further research carried out by Laurel Trainor and colleagues has attempted to dissociate changes in the brain's response to the perception of musical tones based on musical training versus biological maturation. In one study (Fujioka, Ross, Kakigi, Pantev, & Trainor, 2006), the researchers measured neural responses in two groups of children (aged 4 to 6 years old) over the course of four experimental sessions spanning a year. One group of participants was not receiving any musical training whereas the other was involved in Suzuki violin training (a technique that focuses particularly on auditory learning of performance). Neural responses were measured using magnetoencephalography, which measures synaptic activity like EEG but by measuring changes in the associated magnetic field rather than electrical changes. Responses to auditory stimuli increased to some degree in all participants over time due to maturation of the auditory cortex. However, those with Suzuki training developed a selective response to violin tones as opposed to noise bursts. Thus it appears that musical training encourages both note and timbre-specific audio-motor associations, beyond what one might gain through the normal course of development.

Over time, practice strategies that develop aural skills may enhance perception/action associations more so than practice strategies that do not engage aural skills. Seppänen, Brattico, and Tervaniemi (2007) studied auditory processing in two groups of musicians. One group, referred to as the 'aural group,' reported that they improvised at least once a day, played by ear often or quite often, and practiced by listening to recordings of music they were learning. By contrast, the second group reported that they seldom or never engaged in these activities. Participants were presented with stimulus pairs that could be single tones differing with respect to a basic sound feature (e.g., frequency, intensity) or could be pairs of melodies differing with respect to a more complex sequential characteristic (e.g., a melodic interval between two pitches). The researchers measured participants' brain activity using EEG (see chapter 4) while participants listened passively (i.e., without any listening instructions or concurrent task) to these pairs. In most respects, these groups of listeners performed similarly. However, event related potentials to changes in a melodic interval were more rapid for the 'aural' group than the comparison group. Thus, the use of *aural* skills (such as listening to aural models during practice, and engaging in improvisational activities associated with 'informal practice') may facilitate the brain's processing of subtle, higher-order information in musical sequences as opposed to basic features of sound.

Coda

The growing interest in the research on music practice reflects a departure from the notion that music achievement is mainly determined by inborn talent, a 'gift' endowed to the 'lucky' few. Although the contribution of genetics is not ignored in the deliberate practice model (e.g., see Lehmann & Ericsson, 1997) and exceptional ability may not be entirely due to training (Winner, 2000), the role of genetic factors may have been overemphasized in academic and popular assumptions about musical achievement. A growing number of studies suggest that it may be the so-called 'nontalent components' – including motivation, effort, and persistence – that distinguish the highest achievers in most fields (Amabile, 2001). As a whole, the literature on the acquisition of musical expertise gives us reason to question the popular notion that 'some people are just born with it.'

11 The psychology of music performance

In a grand concert hall, the lights dim, conversations hush, and the rustling of programs quickly diminishes to an attentive silence. The performer strides across the stage to the sound of spirited but controlled applause. She takes her seat at the piano and lifts her hands in the air for a moment in a dramatic gesture – before filling the hall with cascading mellifluous tones. The music is emotional, yet subdued. The tempo varies slightly to yield just the amount of emotional expressivity that is characteristic of the historical period of the piece. The audience sits silently and attentively, almost motionless until the end of the performance, whereupon they erupt into another round of enthusiastic but restrained applause.

This scenario conveys a common conceptualization of 'music performance.' Both historically and anthropologically, however, the account above is an exception to the rule. In most places and times, music performance is a collective activity, shared by all persons who are simultaneously perceivers and producers of music. Most music of the world is also an invitation for participants to dance and move to the music, as we will discuss further in chapter 15. One immediately thinks of the use of music in traditional cultures in which music performance serves the purpose of bringing together a village or army. Or compare the scenario described above to a typical rock and roll concert, in which not only standing up and moving to the music, but singing along with the musicians is necessary for the performance to be considered a 'success.' Consider also more intimate uses of music performance. The musicologist Ellen Dissanayake (2000) has proposed that the most important use of music occurs during interactions between mother and child when mothers sing lullabies and play songs (as we discussed in chapter 8).

For these reasons, the current chapter concerns music performance in all its guises: expert and novice, collective and solo, ad hoc and pre-planned. Nevertheless, because the literature to date has typically followed the model of solo, expert, concert performance, much of the research on performance as a whole assumes this model. As the examples above show, performance is also a richly social activity. The social aspects of performance – the group dynamics of musical ensembles, the social rules of performance and concert audience, the facilitating and inhibiting influence of the presence of an audience (including the topic of music performance anxiety), and the effects of the performer on the audience – will be explored in the next chapter.

The plan of the current chapter is to focus on the cognitive and motoric processes that contribute to an individual's performance of music. We trace the progress of a performance in three stages: the formation of a performance plan by using short- and long-term memory, controlling motor actions involved in performance, and monitoring the outcome of one's actions. We follow this discussion with a brief summary of neuropsychological research on performance. For background to this chapter, we drew on extensive reviews of music performance, specifically those written by Alf Gabrielsson (1999, 2003) and Caroline Palmer (1997).

The role of memory in performance

Music performance is a kind of memory task. The actions that give rise to musical events relate to a string of musical ideas from the performer's mind. These ideas, even when entirely spontaneous (which, as we will point out, may be very rare), need to be maintained in memory for some period of time prior to performance. Having said that, the nature of memory demands can vary widely. The soloist described at the beginning of the chapter has an onerous memory task. Such performers usually memorize long and complex pieces written down in a standard notation over hours of practice. Some conductors are known for possessing vast stores of musical memories. The conductor Toscanini was famously reported to reassure a bassoonist with a broken key that the particular key would not be used in that evening's concert, drawing on a precise memory not only for the music of that evening but for the relationship of the music to the requirements on a single instrument in the orchestra. By contrast, someone improvising a solo may draw on memory in a more limited fashion. Nevertheless, every performance involves some kind of memory.

Different forms and uses of memory

The structure of memory constitutes perhaps the most central topic to cognitive psychology. This research suggests that the role of memory may be more far-reaching than is typically assumed. Memories can be stored over the long term or the short term. Memories can be consciously accessible at the current moment or may not be accessible to consciousness.

The classical performer described above likely engages in what would be described as *procedural long-term memory*. Long-term memories are formed when we 'store' information in our minds for some period of time after initial exposure. Typically when people talk about memory, they are talking about long-term memory. The notion of procedural memory, however, can be less familiar. Procedural memory involves memories that are linked to motor actions. Furthermore, we are typically not conscious of the information we are retrieving from memory when we engage in a procedural memory task. An example related to piano performance is touch typing. If you are able to touch type, try to recall what finger is used to press the 'e' key. It is likely that this task was more difficult for you than typing the word 'the,' and more to the point it is likely that you recalled this letter by imagining your fingers striking the keys. Similarly, performers (of the piano or other instruments including voice) do not need to know consciously the identity of each note they perform. They may even disengage conscious thought from the performance, a possibility supported by recent evidence that the frontal lobes effectively 'shut down' during performances from memory (Parsons, Sergent, Hodges, & Fox, 2005).

The example of Toscanini is somewhat different in that the memory is *explicit* rather than procedural. Explicit memories are accessible to conscious thought and are often easy to verbalize. Conductors rely on conscious awareness of the different instruments that come together to create music, and in this way what they remember may differ from the memory demands on the performers.

The fact that procedural memories are not accessible to consciousness does not imply that they are somehow 'easier' than explicit memories, even though we are impressed by memory feats such as that of Toscanini. Musicians spend long, arduous hours, committing music to procedural memory, and it is apparently important that these memories become 'unconscious' so that the performer can devote conscious thought to other factors like expressivity and the audience's response. As discussed in chapter 10, extensive research by Roger Chaffin and colleagues (e.g., Chaffin & Imreh, 2002; Chaffin et al., 2009) documents the kind of strategies performers use to encode music in long-term memory (see also Aiello, 2001; Williamon & Valentine, 2002). It should be noted that performers of music not associated with notation (e.g., rock musicians) still draw on a long-term procedural representation; all that differs is the initial form of the information to be memorized.

What about circumstances in which one does not have to rely on long-term memory, such as a performer who uses notation or a performer who improvises 'spontaneously'? Such performers (in fact all performers) use a type of memory called *short-term working memory*, which refers to the maintenance of information in consciousness. Although the musical notes are provided in front of them, musicians must maintain this information in consciousness for some period of time prior to their performance, and may even maintain the information in memory for some time afterwards (as discussed later). More to the point, classic research on the role of eye movements suggests that performers do not look at the notes they are immediately about to play, but instead read music in an anticipatory manner, attending instead to notes in the future. This can be demonstrated by asking someone to sight-read music at a moderate pace, taking away the notated music, and counting the number of notes the person is able to continue to play without the music. The average 'eye-hand span' for instrumentalists is about five or six notes

(Sloboda, 1974). Thus, at any point in time, the performer probably produces notes that have persisted in memory for at least a few milliseconds and that are not providing immediate sensory data.

Improvised performances provide the least obvious connection with memory, in that these performances, at first glance, seem to be completely spontaneous. This problem is compounded by the fact that the cognitive bases of improvisation are not as well understood as are the bases for performance from notation (Ashley, 2009). However, anecdotal observations and analyses of improvised solos do suggest some use of memory-based processes. First, most practicing jazz musicians will dispute the notion that jazz solos are entirely spontaneous, with the simple fact that jazz musicians *practice* certain characteristics of solos (as noted in chapter 10). Although no two solos are the same, solos tend to result from permutations of musical motifs that are well practiced. In this way, jazz solos mirror language. Although most statements are creatively novel when taken as a whole, they typically consist of shorter phrases that we speak quite often. A second way in which jazz solos involve memory stems from the fact that they incorporate notes that abide by a certain rule system, referred to as a *grammar*, that fits within the genre. Jazz musicians thus rely on *schematic* information, which is commonly considered to reside in memory (e.g., Johnson-Laird, 2002).

Thus it can be said that all performers rely to a great degree on short-term working memory (henceforth working memory), even when performing from notation. Given its importance, we now turn to research that focuses specifically on the role of working memory in performance.

Working memory and planning of serial order

One of the most influential papers concerning the cognitive bases of production was a paper by Karl Lashley entitled *The Problem of Serial Order in Behavior* (1951). At the time, Lashley was lashing out against behaviorists (e.g., B. F. Skinner) who had suggested that retrieval during sequence production resulted from a *serial chain* of events in memory. Serial chaining approaches suggest that people associate event ♯1 (e.g., the note C) with event ♯2 (the note D) and so on when retrieving events during a performance. Lashley claimed that such theories are untenable. First, people can produce sequences much more quickly than a chaining approach would predict. Pianists can produce sequences of notes in which the time interval between notes is as rapid as 16 notes per second (about 62 ms per note); the timing of neural events, given refractory periods in neural impulses, could not support such rapidity.

Lashley's second point is more germane to this chapter. He also pointed to a certain kind of error known as the *serial ordering error*, as evidence against serial chaining. A serial ordering error is one in which people misplace the order of a particular event while maintaining the appropriate content. Here is an example from speech, heard once by one of your authors (PQP):

Spoken:

'We are here to celebrate the wedding of Kelley and Sheith . . .'

Intended:

'We are here to celebrate the wedding of Shelley and Keith . . .'

In this instance, the priest at a wedding swapped the position of two speech sounds (phonemes), the 'Sh' sound from 'Shelley,' and the 'K' sound from 'Keith.' Thus in the spoken example, the priest produced all the correct phonemes but misplaced the order of two of them.

Why do serial ordering errors contradict serial chaining approaches? The reason is simple: Serial ordering errors usually span multiple positions. Take the example above. The sounds that get swapped do not come from adjacent phonemes. Instead the sounds travel a distance of approximately seven phonemes away (? – e – l – y – a – n – d – ?). Instead of serial chaining, the error above (and other errors like it) shows how speakers relate sounds to the structure of language. It is significant that the priest swapped two sounds that were associated with the beginnings of proper names. The psycholinguist Stephanie Shattuck-Hufnagel (1979) suggests that when we plan speech sequences we use a *slot-filler* mechanism, whereby the appropriate content (e.g., phonemes) is applied to slots associated with the structure of language.

Serial ordering errors occur in music performance as well. Research to date has focused on piano performance and it appears as though 'slips of the finger' do not merely result from 'missing' the intended key and instead hitting the one next to it. Just like errors in speech, musical errors seem to reflect mental associations of content (the notes) with the global structure of music. A typical example of musical error is shown in Figure 11.1 (from Palmer & Pfordresher, 2003). Note that the performer does not play an error that involves hitting an adjacent key to the intended note (D), which would have been D♯ or C♯. Instead, the error involves the inappropriate repetition of an earlier note, G, which was intended for position 1. The error event thus travels a distance of 2 positions (from the target to error location). Such an error is

Figure 11.1 An example of a serial ordering error in music performance (Palmer & Pfordresher, 2003, figure 8).

Source: Reprinted with permission from the authors and the American Psychological Association.

also referred to as a *perseveratory* error, as opposed to *anticipatory* errors that involve playing a note in advance of its correct location. One difference between speech and music is that musical errors rarely 'trade' places as in the example above (referred to as an 'exchange' error).

Some of the pioneering research on this topic suggested that serial ordering errors in music are constrained by the structure of music. That is, the degree to which notes 'travel' to incorrect locations in production depends on how the performer conceptualizes the structure of the music. For instance, Palmer and van de Sande (1993) found that in piano performances for which each hand plays an independent melodic line (*polyphonic* music), notes from within the right hand's melody tend not to replace notes from within the left hand's melody. In later work Palmer and van de Sande (1995) demonstrated that errors only travel within a single musical phrase (analogous to a clause in speech). The implication of both findings is that at any point in time, the performer has on her mind a certain subset of the notes in a melody – for instance, the notes within a phrase – and that notes may be further encoded with respect to a particular melodic line out of which they do not travel. Figure 11.2 illustrates these kinds of boundaries. The figure shows a pianist retrieving from memory while performing Bach's two-part invention in A minor (used in Palmer & van de Sande, 1995). The pianist is currently producing the second phrase, hence phrase #1 is shown in gray. Within the second phrase the two parts are conceptualized as separate and are thus shown within different ovals.

A broader implication of these findings is that although a musician must have the entire piece stored in memory, the process of retrieval during performance does not result in the performer retrieving the entire piece. Nor does the performer retrieve the piece one note (or one chord) at a time, as might follow from a serial chaining approach. Instead, memory retrieval is *incremental* in nature. At a given point in time the performer holds some, but not all, notes from a performance in working memory (cf. Baddeley, 1986; Miller, 1956). The notion of incrementality, like analyses of serial ordering errors, originated in speech production research as a way of accounting for the fact that people seem to plan future utterances (e.g., words) as they continue to produce the current utterance (e.g., Kempen & Hoenkamp, 1987). This is particularly notable when one is speaking a second language.

The depiction shown in Figure 11.2 suggests that the performer may have equal accessibility to all the notes within the second phrase, that these notes come as a block or 'chunk.' Indeed working memory theorists have often invoked the idea that items or events may all be treated as a common 'chunk' of information. However, such a proposal seems untenable as an account for how people retrieve events in a sequence. It would predict that performers produce any note from within a phrase with equal probability. Such a situation would lead to very low accuracy, and performers are typically highly accurate (less than 5 percent of all performed notes are usually errors [e.g., Palmer & Pfordresher, 2003]). Furthermore, serial ordering errors often

Figure 11.2 A schematic illustration of structural boundaries in a performance plan.

Source: Notation from Palmer and van de Sande (1995, figure 2, p. 951). Copyright 1995 American Psychological Association. Used with permission of first author.

reveal some notes within a phrase to be more likely error sources than others. Thus, an incremental approach has to account for why some notes within a phrase are more accessible than others.

Such an account has been developed in recent years by one of us (PQP) and Caroline Palmer (Palmer & Pfordresher, 2003). We have developed a model we refer to as the *range model* because it accounts for the range of events that are accessible in working memory as one plans and produces a melody. The model also accounts for degrees of accessibility across events. Although the model was described mathematically in the original papers, a verbal description will suffice here. It presupposes that one maintains some knowledge of the current position. The model is also limited in that it does not account for the specific notes associated with positions but rather simply accounts for how mentally accessible each position is (in the memory jargon, the model includes 'order' but not 'item' information).

Within these constraints, the range model proposes that accessibility of notes within a phrase is governed by two components. First, notes that are closer to your current position are more accessible than notes that are far away. For instance, if the current note is note ♯5, then notes at positions 6 and 4 will be more accessible than notes at positions 8 and 2 (see Figure 11.3, lower left). Second, the model predicts that notes are more accessible if they are associated with a similar 'metrical accent' to other notes. Thus, when one plans to produce a note or chord on a 'weak' beat, notes and chords on other weak beats are more accessible in the brain than are notes and chords on strong beats (see Figure 11.3, lower right). The combined predictions of these two components are shown in the top plot of Figure 11.3.

The top panel illustrates what the performer may have on her mind during a performance (an elaboration on the callout shown in Figure 11.2). The grid of Xs below the music illustrates the pattern of metrical accents at each position (more Xs = more strongly accented; see chapter 6 for more detail). The current position (highlighted by the rectangle superimposed on the notation) is a strong metrical position. The bars in the graph above the notation

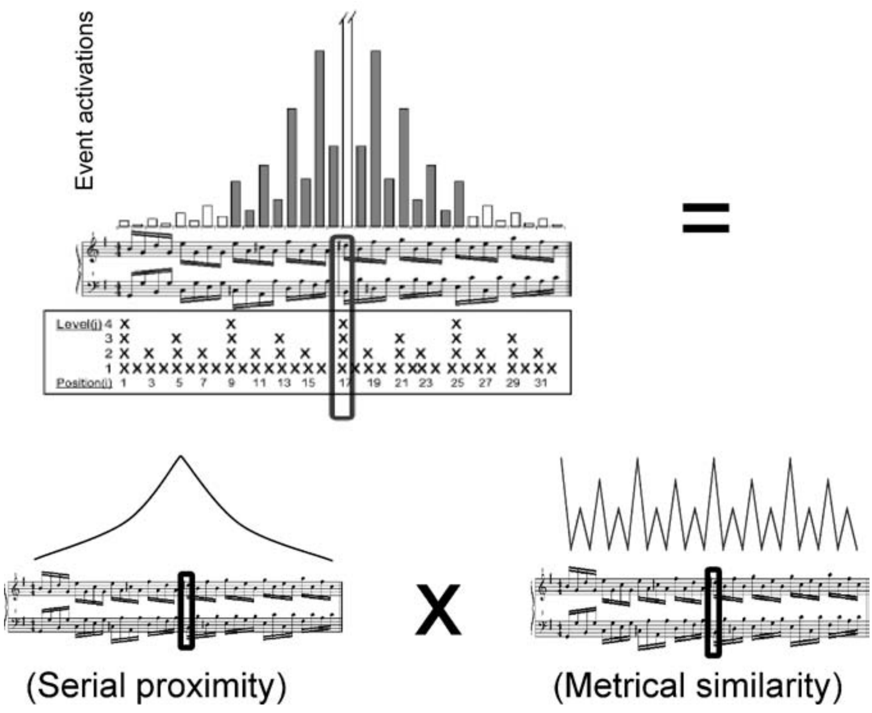


Figure 11.3 Predictions of the range model (top) along with predictions of the serial component (lower left) and metrical component (lower right). Adapted from Palmer and Pfordresher (2003, figures 3 and 5).

Source: Adapted with permission from the authors and the American Psychological Association.

illustrate how accessible or active each note is – in other words, how likely it is that each note will be performed. The current note is most highly active and thus is the most likely to be performed (note that the bar extends above the plot). However, other positions are also active by virtue of both their proximity to the current event, and how similar their position is to the current position with respect to metrical accent. What this means for the performer is that if he or she makes an error when intending to perform the D♯-note in the right hand (at the highlighted position), that error is more likely to be F♯ (2 events in the future) or D-natural (2 events in the past) than it is to be either B or A (1 event in the future or past, respectively). These predictions have been tested and verified across many studies of piano performance from memory or from notation, at different tempos (speeds), and for both adult and child pianists (reviewed in Palmer & Pfordresher, 2003).

The research on the role of working memory in performance is quite recent, and many questions remain. Two particularly compelling questions include how well these findings generalize to performance of instruments other than the piano and to singing, and how they generalize to improvised performance. We now consider what happens after a performer retrieves the notes that make up music and puts those notes into action. How do performers control the actions that are used to bring a performance to life?

Controlling movements during performance

Music performance is most obviously a form of motor control; the fine-grained control of rapid movements that is exhibited by expert performers is impressive indeed. We will consider movements from two perspectives. First, we will consider the way movements are controlled with respect to timing. Second, we will consider how performers control movements in a way that leads to an appropriately 'tuned' outcome with respect to pitch. This second issue is not a concern for pianists, given the fixed nature of pitches, but can be daunting for many other instrumentalists (e.g., French horn players) and singers. In discussing these issues we will consider both the basic issue of how it is that performers can control movements so that the produced music matches a communicative intent and the way in which movements are modulated to give the performance aesthetically appealing 'expressive' nuances.

Controlling the timing of movements

As discussed in chapter 6, the best cue for rhythm is the so-called *inter-onset interval*, the time that elapses between successive note beginnings. Research on performance has traditionally focused on the timing of movements that are directly related to inter-onset intervals. For piano, these movement points comprise the times at which the piano key hits its key bed. Such measures of timing have been used because they are musically relevant, but also for convenience, as the measurements can be derived from the electronic output of a keyboard (e.g., one that uses MIDI). However, because this kind of analysis includes only selected discrete time points a lot of information is lost.

More recently, researchers in music performance have harnessed a new form of technology called *motion capture*, which involves the analysis of movement in real time. Motion capture can address movement in three spatial dimensions, and can involve measurements of movement position as well as higher-order variables (e.g., velocity, acceleration). Figure 11.4 shows an example of video-based motion capture, in which movement analyses are based on extraction of image boundaries and the movement of those boundaries across time (Castellano, Mortillaro, Camurri, Volpe, & Scherer, 2008).

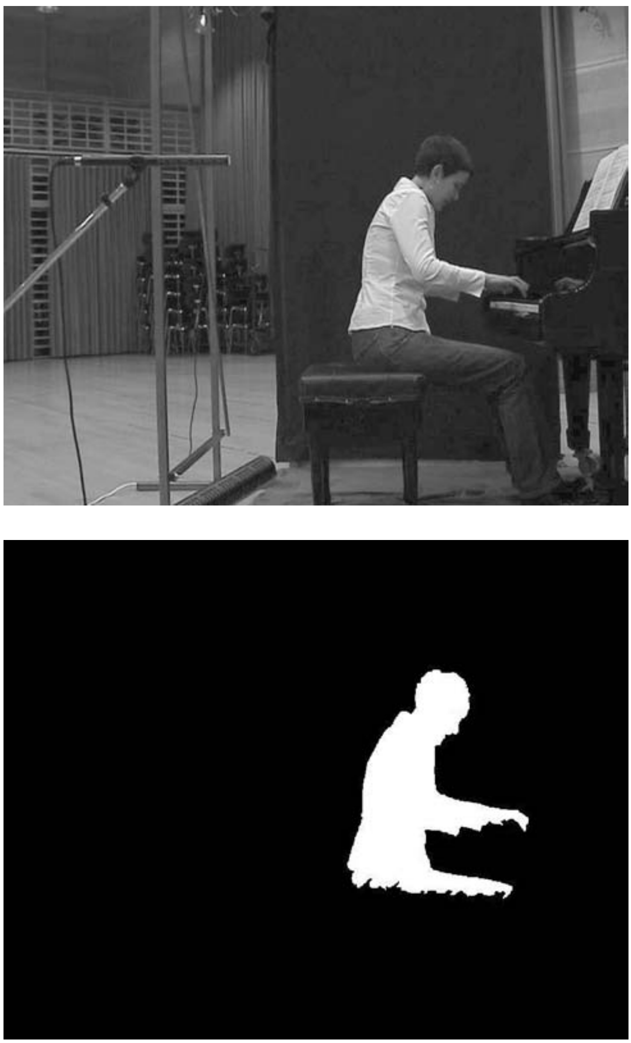


Figure 11.4 Video motion analysis of expressive piano performance (Castellano et al., 2008, figure 2). The first two photographs show a single frame from the original recording (top) and the extracted pianist's silhouette (bottom). Bright white areas in the third photograph highlight regions where movement is detected in a single frame following motion analysis.

Source: Reprinted with permission from the University of California Press.

Other forms of motion capture involve attaching small markers to the performer's joints; analyses then focus on the movement of these markers (e.g., Loehr & Palmer, 2007).

Maintaining regularity in timing

A major task of the performer is to produce inter-onset timing in a way that successfully maintains the sense of a regular underlying 'beat,' which gives the listener a regular sense of how rapidly musical events are proceeding (the 'tempo'). How do performers do this, and how well do they achieve it?

Imagine a melody comprising a sequence of notes with equivalent durations, and no expressive nuances to the timing of these notes. Such a performance (probably boring to hear) would entail perfectly *isochronous* timing (i.e., timing with equivalent durations). Oddly enough, nobody can do this precisely. Inter-onset timing always varies from event to event, even when one intends to perform isochronously. Good performers may minimize this variability when they wish to, but no human can eliminate it. Even electronic drum machines typically have some variability (Levitin, personal communication).

Why do these inaccuracies occur? The dominant view at present comes from the work of the psychologists Alan Wing and Alan Kristofferson (1973). Their original research addressed isochronous finger tapping (tapping the same beat repeatedly), but is applicable to piano performance as well. Their model, referred to as the 'two-level timing model,' suggests that timing inaccuracies are the sum of two components, and both are inaccurate. The first component is an internal time-keeper, or 'clock' (see chapter 6 for further discussion). Because nobody is perfect, the internal clock is inaccurate. In addition, the *production* of time intervals requires the use of muscles, and the precision with which we use muscles also leads to inaccuracies referred to as 'peripheral variability'. The types of inaccuracies that occur involve the addition of random variability to the time one attempts to keep (e.g., the tempo). This is in the spirit of 'clock-counter' models discussed in chapter 6. Thus while someone attempts to keep the time of a 500 ms interval, variability might cause this interval to vary from 490 ms to 410 ms from moment to moment.

Typically, musical training focuses on reducing variability at both the clock and peripheral levels. Percussionists arguably focus the most on reducing internal clock variability in addition to peripheral variability, whereas other instrumentalists focus more on peripheral variability. One intriguing prediction from this model is that when a musician switches to a new instrument that requires the use of different muscles (e.g., clarinet to drums), internal clock variability should remain consistent but peripheral variability should increase.

This model has been very influential and is simple in that it includes only two components. However, simplicity can come with a cost and there are at least two respects in which this model's limitations may be too great. One potential oversimplification is that the separation of clock and motor components implies that the 'clock' is common to *any* motor task, regardless of the muscles used. There is mixed evidence for this claim, Some evidence from studies comparing drawing and tapping suggests the use of different internal clocks for each task (Robertson et al., 1999). At the same time, other evidence from recent experiments using motion capture suggests that pianists may conceptualize the timing of key presses (notes) in a way that is independent of the biomechanical constraints regulating movements (Loehr & Palmer, 2007), which is consistent with the idea of separable clock and motor components.

Lawful variations in timing: rhythms

In chapter 6 we discussed the many ways in which a time pattern may be considered 'rhythmic.' Here we focus on the production of rhythms rather than on the basic factors involved in rhythm, focusing in particular on the role of rhythmic complexity. The ability of performers to reproduce rhythms is determined in part by how complex the rhythm is. A simple rhythm – one comprising proportionally related intervals that form reliable patterns – will be easily learned and reproduced, but not so for a complex rhythm. A famous laboratory example was reported by Povel (1981). He found that when people try to reproduce rhythms consisting of complex ratios, their produced rhythms 'drifted' toward the closest simple ratio (e.g., a 2 to 1 ratio).

Complex rhythms do not exist simply in the lab. A famous example of a highly complex classical rhythm is found in parts of Igor Stravinsky's ballet, *The Rite of Spring*, the music to which a young woman dances to her death. The music presents a very clear rhythmic pattern; however, it is hard to extract any sense of regularity from this pattern, because in fact it is a faithful rendering of tachycardia (rapid heartbeat). As a result, reproducing this rhythm is quite difficult to play and was initially off-putting to listeners as well.

Of course, there are also cultural factors at play. The rhythmic pattern shown in Figure 11.5 is ubiquitous in West African drumming, and can be easily reproduced there, but is very difficult for most European-descended listeners to learn and reproduce even though all the time intervals are related by a 2:1 ratio with each other.

Expressive variations in timing

'Lawful' variations in timing alone, however, would not make for the most expressive performance. 'Expressive timing' includes more obvious deviations from the isochronous beat (*ritardando, accelerando*). However, more nuanced performances also employ subtler, incremental deviations from the beat. For instance, expressive departures from strict metronomic time in the opening arpeggios of Debussy's *First Arabesque* can create a feeling of resistance during the upward swing of the arpeggio, a barely perceptible pause at the very top of the arpeggio, and a sense of release in the downward fall. In the music of J. S. Bach, these departures from isochronous time may be more subtle, a barely perceptible delay between phrases. These deviations from the strict metronomic beat are one of the main sources of musical expressivity and 'expressive timing' probably constitutes the most well-studied aspect of musical performance.

Its tradition extends back at least to the detailed analyses of performance carried out at the University of Iowa by Carl Seashore and colleagues (Seashore, 1938/1967) in some of the earliest empirical research on the psychology of music. In their lab, performances were analyzed and plotted on 'performance scores' that displayed variations in pitch and intensity in real

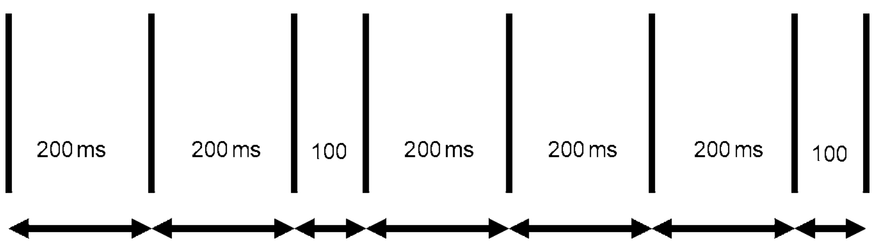


Figure 11.5 Time pattern common in West African drumming.

time. Modern researchers have incorporated simplified versions of these representations by plotting expressive patterns above music notation, so that expressive nuances associated with particular notes align the way that the note is written. Figure 11.6 shows a hypothetical example that is consistent with current literature.

Research in Seashore's pioneering lab, and conducted subsequently, revealed that performers systematically deviate from the instructions listed in musical notation, as shown in Figure 11.6. Suppose a performer is playing a melody at a tempo of 100 beats per minute in 4/4 time, which means that each quarter note should be held for 600 ms. A performer may choose to lengthen one note ever so slightly, say to 660 ms, based on her expressive intentions, or may choose to shorten it. It is important to note that these deviations from regularity in timing are not based on the same factors as those discussed earlier with respect to the two-level timing model (Wing & Kristofferson, 1973). Expressive deviations are linked to intentions of the performer, whereas the variability accounted for by the two-level model reflect the inherent noise in the system – the kind of thing a performer wants to 'reduce.'

A logical question regarding expressive timing is how idiosyncratically it is assigned. Seashore and colleagues showed that individuals can be very consistent in their expressive timing. This finding has often been replicated.

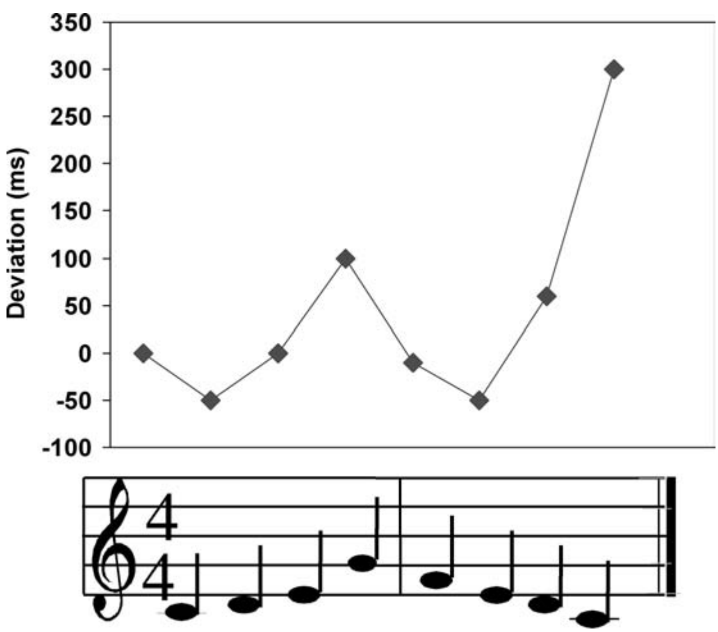


Figure 11.6 Hypothetical expressive timing of a simple melody. The music notation (below) prescribes exact inter-onset timing, but the pattern of produced inter-onset intervals (graph above) shows systematic deviations from exact regularity. Copyright © Peter Pfordresher.

However, how consistent is one performer from the next? Certainly, individuals have varying interpretations, but some common principles hold across performers. Perhaps the most highly generalizable principle is *phrase final lengthening*. Put simply, people tend to slow down toward the end of a musical 'group.' In addition, people tend to speed up at the beginning of a phrase. The idea is similar to language. People tend to pause and to slow down when they reach the end of a clause, such as in:

> Mary went to the store (pause), I went to the bank

Similarly, there are phrases in music, as discussed earlier. For instance, the first two lines of the tune 'Mary Had a Little Lamb':

> Mary had a little lamb, little lamb, little lamb
> Mary had a little lamb whose fleece was white as snow

would constitute two musical phrases, the first a 'question,' the second an 'answer.' When performing these two lines, the performer would be likely to time each inter-onset interval in a manner that matches the phrases.

The timing pattern shown in Figure 11.6 illustrates such a prototypical pattern. The swooping lines above the notation indicate how much each note deviates from what its length would be if the person were performing exactly as the notation dictates. A 'mechanical' or 'deadpan' performance would result in all performed notes matching a length of 0. Note that this is not the same thing as plotting note durations, which could introduce variations in inter-onset intervals that are determined by the notation. In this hypothetical performance the performer begins each phrase metronomically (a deviation of 0) then speeds up to about the middle of the phrase, and slows down to the end of it. Note also that the performer slows down more after the second 'answer' phrase than after the first 'question' phrase.

Although performers do vary one from the other, there is evidence for an overall tendency to time performances in this way, with more complexities arising for more complicated pieces. Todd (1985) proposed a mathematical model that predicts the degree of slowing a performer exhibits to the amount of 'closure' associated with a phrase. Degree of closure in his model was predicted by principles from the theory of Lerdahl and Jackendoff (1983), discussed in chapter 7. Though the predictions of this theory do not perfectly fit every situation (see Repp, 1991 for detailed analyses of a number of famous performances), it provides a starting point. Moreover, some evidence suggests that people can better imitate a performance with 'typical' (i.e., Todd-like) timing than an 'atypical' performance (Clarke & Baker-Short, 1987; but see Repp, 2000 for a slightly different interpretation). Thus, performers can more readily imitate performances of others that conform to standard conventions.

Thus, performers use variations in timing to express musical ideas. In so doing the performer gives the listener an idea about his or her interpretation of the music, and 'gives life' to the performance. Some have suggested that the lively nature of expressive performance comes about because the variations in timing are reminiscent of patterns of physical motion. Expression thus helps propel the music forward. Anders Friberg and Johan Sundberg (1999) tested a model of expressive timing based on an analogy with the way in which runners stop themselves after the fall phase of a step. The resulting curves are similar to those predicted by Todd's (1985) model, though based on slightly different principles. Despite the intuitive appeal of the music-as-motion analogy, this view has detractors. A specific example is a paper by Henkjan Honing (2005) who argued that expressive nuances may be better predicted by a theory based on the performer's need to follow the rhythmic pattern than on patterns typical of large-scale physical motion (like running and dancing).

Other aspects of timing may be used to communicate the performer's interpretation. Consider, for instance, timing among many notes that are notated as sounding simultaneously. Performers typically don't perform all these notes at once, and these slight deviations may be linked to expressive intentions. Specifically, performers appear to use *voice leading*, which occurs when one note is produced slightly before the others. Voice leading has been studied in solo piano performance (as has expressive timing), in which context one can make the easiest case for a link between voice leading and intentions. Caroline Palmer (1989) has reported evidence suggesting that voice leading is used to highlight which of the pitches that are notated as being simultaneous is most important, i.e., which one constitutes the melody. In other words, voice leading helps one part 'stand out' to the listener. In addition, there is evidence from perception that asynchronous events are less likely to sound as though they belong to the same Gestalt group than are synchronous events (Bregman, 1990). Some have suggested that voice leading may be linked more to the velocity with which certain fingers strike the keys (related to the loudness of notes) than to timing per se (Goebl, 2001), though this alternative account still links voice leading to expressive intentions.

Additional expressive information may be communicated by performers visually through movement (as discussed further in chapter 13). Jane Davidson (1993) has recorded performers' movements while playing the piano, focusing on movements of the trunk, arms, and head. Her data show that performances associated with more movement are perceived as being more expressive than those with less movement (e.g., when a performer's movements are constrained). This perceived expressivity was found even when people saw point-light displays instead of the entire film. In point-light displays, the viewer sees little dots of lights positioned at individual joints. Thus a pattern of moving dots is all the viewer has to reconstruct the underlying motion, although it is certainly more aesthetically rewarding to see the whole performer! More recently, Davidson (2005) has argued that expressive movements are hierarchically related to a common 'center of movement,' which for a pianist may be the waist. Recent results from a study using motion capture furthermore suggest that head movements play a primary role in the communication of emotion for piano performance (Castellano et al., 2008).

Controlling the tuning of movements

So far much of the literature reviewed on the research on music performance uses the piano as the main model. This focus has much to do with the ease of measuring timing in piano performances. Current technology allows one to record the exact time at which a piano key contacts the key bed, whereas calculating timing for other instruments typically requires more complex analyses based on the acoustic waveform. However, examining the piano is limited with respect to the next factor in motor control during performance, the 'tuning' of a performance. We now focus more on the voice and other instruments as we turn our attention now to tuning.

Tuning refers to the precision with which a performer generates musical events that match intended pitch categories. Consider the prototypical 'bad singer.' Such individuals have been found to sing with appropriate timing but may sing the wrong notes entirely (Dalla Bella, Giguère, & Peretz, 2007). This is clearly a case of poor 'tuning.' In other respects, tuning may be used expressively as in the case of vibrato or slight deviations from a pitch class (Morrison & Fyk, 2002). In contrast to research on timing, research on tuning concerns instruments other than the piano, which is tuned prior to performance. The most critical instruments from the standpoint of tuning are unfretted stringed instruments (e.g., the violin) and the human voice.

Mechanisms for tuning

We focus on singing, as it is a ubiquitous musical behavior and poses perhaps the most difficult task with respect to tuning. Singing in tune is accomplished by varying the tension and length of the vocal folds (or cords) in one's larynx. One cannot see these muscles without the aid of a laryngoscope (which involves inserting a tube down one's throat); furthermore, it is impossible to accurately gauge the muscle tension that is generated, rendering both visual and proprioceptive feedback useless. By contrast, violinists can observe the position of their fingers (though experts do not do this), and trumpeters can feel the tension of their lips (embouchure) to control tuning.

Vocal production involves three basic processes: respiration, phonation, and articulation (see Sundberg, 1987 for an extensive review). Respiration refers to the control of breath (inspiration, breathing in, and expiration, breathing out). You may have been told about the virtues of singing or speaking 'from the diaphragm.' This statement refers to the use of lower abdominal muscles to push air up out of the lungs. The voice works best when a strong column of air is sent up through the trachea (the air tube through your throat). Many untrained singers, by contrast, sing by tightening their chest muscles, which results in a more strained sound.

Phonation contributes the most to tuning. The process is quite complex and we will not offer the complete details here. Inside your larynx are a set of muscles and bones that govern two flaps of skin known as the 'vocal folds' (or vocal cords). Vocal folds can be brought together (*adduction*) or held apart (*abduction*). The folds are held apart when we whisper or breathe, and a pitch-like quality results when folds are held together (adduction). During adduction, air from the lungs causes the vocal folds to vibrate back and forth, and the resulting frequencies of vibration generate pitch. Phonation refers to the generation of this pitch quality. The length, tension, and massiveness of vocal folds are all varied when we produce different pitches. In general, high pitches are created when the folds are tenser, less massive, or (counter-intuitively) longer. Falsetto voice, for instance, is produced when the vocal folds are stretched out and thinner.

The final process in vocal production, articulation, arguably is crucial for the communication of words, but not quite so much for tuning. This stage involves the manipulation of the tongue, lips, and jaw in order to articulate speech sounds. More generally, these alterations can influence the sound quality (timbre) and can make the sound of one's voice rich and full (if sound is directed out of the mouth) or nasal (if sound is directed out of the nose).

Figure 11.7 shows the effect of articulating different sounds on the power spectrum of the voice (see chapter 2 for an explanation of these kinds of plots). These plots show someone singing three syllables at the same pitch. Although fundamental frequency is the same in each plot, one can see that

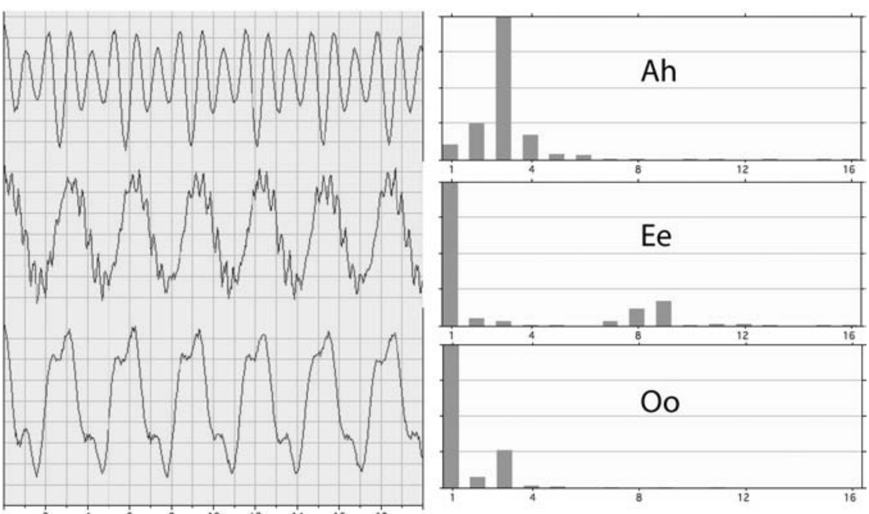


Figure 11.7 Waveforms (left) and power spectrum plots (right) for three syllables sung at about 325 Hz (close to an E above the middle C).

Source: Figure created by Deedré Thomas Cousins and composited by Rod Nave at the Department of Physics and Astronomy at Georgia State University, used with permission.

the spectral structure differs considerably. These changes to spectral structure are critical for the intelligibility of singing with words, in that it is the spectral changes across time that allow the listener to perceive and comprehend spoken words or a sung lyric such as 'hallelujah!'. The commonality in fundamental frequency across the syllables can be seen in the waveforms (left of Figure 11.7) in that each shows a repeating pattern that recurs about 6.5 times in the time-span shown.

Individual differences in the tuning of actions

Of course, once we mention singing, individual differences leap to mind as an important issue. We've all had the painful experience of hearing a well-intentioned friend sing a 'melody' that bears little resemblance to the intended tune, and we can all think of professional singers who perform astounding vocal feats. A surprisingly large percentage of people claim to be 'tone deaf.' At least 17 percent by a recent estimate (Cuddy, Balkwill, Peretz, & Holden, 2005) consider themselves 'tone deaf,' and in a survey of 1105 introductory psychology students 59 percent claimed to be unable to imitate melodies by singing (Pfordresher & Brown, 2007). Is incompetence in vocal tuning as prevalent as we think it is?

First of all, it is important to distinguish mistuning of the voice during singing from other vocal dysfunctions that may result in less-than-ideal vocal quality. Many people, for instance, have 'rough' sounding voices. This quality occurs when the vocal folds are not held close enough, and some of the air stream gets through. A typical cause of this problem is from calluses ('nodes') that form on the vocal folds, often after years of abuse (singing too loud). Many rock stars who are apparently able to scream for years with little effect on their voice have learned special screaming techniques (believe it or not) so that they, like a baby, can wail for hours without damage. But the untutored screamer is unfortunately destined for a brief singing career.

The cause of the kind of mistuned singing that we have all experienced is, surprisingly enough, unknown. Some interesting possible answers have emerged only recently, but we are far from fully understanding the phenomenon. An intuitive explanation focuses simply on motor control: Bad singers may be unable to control their vocal folds. However, evidence in favor of this view is lacking, and some reflection can determine why. Even bad singers are typically adequate at using pitch to communicate in language, as in differentiating a question from an answer. Moreover, bad singers may be no less precise in controlling pitch spontaneously than are good singers. The differences between people appear in situations where one has to imitate rather than generate the structure of a sequence, particularly musical sequences.

If motor control doesn't explain poor singing, what does? The common usage of the term 'tone deaf,' as an indicator of poor singing (Sloboda, Wise, & Peretz, 2005), suggests a perceptual basis. Plainly put, inaccurate singers may also be inaccurate listeners. True tone deafness would suggest that inaccurate singers have a reduced ability to hear pitch differences, resulting in a higher *difference threshold* (the minimal pitch difference one can detect) than found in the normal population. Recent highly influential research by Peretz and colleagues has in fact revealed a population with such a deficit, and this population is also less accurate at singing than the rest of the population (e.g., Ayotte, Peretz, & Hyde, 2002). This deficit in pitch perception has been termed *congenital amusia*, suggesting an inherited disorder of music processing.

Thus, there are likely to be at least some inaccurate singers who suffer from a perceptual deficit. But is this true of most bad singers? Probably not. Other research has identified inaccurate singers who are apparently unimpaired when given perceptual tasks (Bradshaw & McHenry, 2005). If bad singers do not suffer from either a motor or a perceptual deficit, what is the basis of their impairment? Other possibilities include poorly formed schemas for motor plans (Welch, 1985) or an inability to link perception with action (Pfordresher & Brown, 2007).

Finally, although inaccurate singing is irritating to hear, it may not be as pervasive as most people think. Dalla Bella and colleagues (Dalla Bella et al., 2007; cf. Davidson, 1994) recently conducted an extensive study involving both field and laboratory recordings of people singing a familiar melody ('Happy Birthday' in French). They found that all but two participants (about 10 percent of the sample) were about as accurate as were professional singers! It should be noted that 'accuracy' here means singing closer to the correct pitch than to its neighbor. It is quite possible to sing 'accurately' but still sound slightly mistuned, which is probably the case for most 'bad' singing.

Monitoring the outcomes of performance

We now turn to a topic often overlooked in accounts of performance: the role of performer as perceiver. We refer to the perception of the results of one's performance as *perceptual feedback*. Perceptual feedback can span many modalities, including audition, vision, haptics (touch), and *proprioception* (the feeling of one's own movement based on feedback from the muscles and joints). In practice, almost all the research on self-perception in performance has focused on auditory feedback, which we focus on here. In this chapter we focus on the role of feedback from one's own actions; in chapter 12 we discuss the role of feedback from other performers and from the audience.

It may be taken as a truism that the performer has to hear what he or she has done in order to continue performing effectively. This makes sense from what has been called an 'error monitoring' account (e.g., Levelt, 1989): People use feedback to gauge whether they have done the right thing. But this 'truism' may not be true, at least not for performances of well-learned melodies. In fact, when pianists perform with the sound turned off, some subtle differences in a performance result but the changes to performance would not constitute 'disruption' per se (Repp, 1999). Singing (Mürbe, Pabst, Hofmann, & Sundberg, 2003) and playing of stringed instruments (Chen, Woollacott, Pologe, & Moore, 2008) may rely more on auditory feedback, in that intonation suffers when auditory feedback is absent (or masked). However, changes to intonation in such cases are not so large as to constitute 'errors' in pitch production, as can be produced by altered (as opposed to absent) auditory feedback.

Though the presence of auditory feedback may not be crucial, it is crucial that auditory feedback – when present – match the sequence of executed actions in a performance. Alterations of auditory feedback can be, though they are not always, devastating to a performance, such that a skilled pianist sounds like a beginner. The most widely studied alteration to auditory feedback is delayed auditory feedback (DAF). First explored in the domain of speech (by Black, 1951; Lee, 1950), DAF simply involves adding a temporal gap between the time at which an action is executed (e.g., a key press) and the time at which the resulting sound begins. It is like playing in a large echoic room (e.g., a cathedral), only in this room you can only hear the echo, not the initial sound of your instrument. (Incidentally, pneumatic pipe organs function a bit like this. If you have ever played one, you may recall a noticeable delay in the production of sound whereas most delays in the response times of musical instruments are not clearly perceptible.)

Unlike the absence of immediate auditory feedback, the influence of DAF on performance is profound. For performers, DAF (for delays around 200 ms and larger) can slow down timing, cause timing variability to increase, and increase error rates. DAF's strongest effects, however, may be on timing of actions more so than whether the correct note is played (Pfordresher, 2003). The effect does not appear to be related to attention; attempts to manipulate attention to sound conducted in unpublished research by one of us (PQP) have not influenced DAF's effect. Rather, the effect of DAF – at least in music – appears to reflect the presence of disrupting rhythmic relationships between perception and action (cf. Howell, Powell, & Khan, 1983).

Another characteristic of auditory feedback that has been investigated more recently is pitch. Interestingly, simply altering every pitch you hear in a random or random-like way while playing piano does not disrupt production (Finney, 1997). So disruption does not result from hearing an artificially induced 'error.' However, hearing altered pitches that form a sequence like the one being played, that does not line up appropriately with the sequence already produced, can be very disruptive. For instance, hearing a melody with the same contour, but having all the upwards intervals in the melody sound when the performer is producing downwards pitch changes on the keyboard, increases errors significantly, though it does not similarly disrupt timing (reviewed in Pfordresher, 2006).

Taken as a whole, these findings lead to an important implication regarding the way in which performers use perceptual feedback. Put simply, performance and perception draw on the same neural resources. Recent neuroscience research has revealed so-called *mirror neurons* that may be responsible for the mental simulation of action (see neuroscience section from chapter 10). According to this view, auditory feedback that presents a similar, though displaced, sequence to the one a musician is trying to perform may be activating mirror neurons at the wrong time, leading to the generation of inappropriately timed action plans.

Many remaining questions exist regarding the use of feedback in performance. Research to date is largely limited to audition, as opposed to other modalities such as vision and touch, and to feedback from oneself rather than from others. The situation is beginning to change. For instance, recent research has investigated the effect of performing with an ensemble in which information from other ensemble members is delayed (Bartlette, Headlam, Bocko, & Velikic, 2006), and the effect of singing in a chorus when other parts are mismatched relative to one's own (Fine, Berry, & Rosner, 2006). Both studies document disruptive effects comparable to those found with self-feedback. Finally, we should not forget the audience, who can be a substantial source of feedback. The social psychological aspects of the presence of an audience are discussed in chapter 12. From a performance point of view, any performer knows that when there is a large and responsive audience, the effect on one's performance can be profoundly enhancing, and that the opposite can occur for small or disinterested audiences. There is much room for future research on this important topic.

Music performance and the brain

In the previous chapter, we reviewed research from the field of cognitive neuroscience on changes in the brain brought about by the practice of music. Here we focus on research concerning the performer 'as is.' It is worth noting at the outset that many of the areas discussed below are active even when one simply imagines music (Halpern, 2003); thus the neural 'performance' of music may be central to musical experience in general.

One glimpse into the way that the brain coordinates various activities during piano performance was given by Sergent and colleagues (Sergent, Zuck, Terriah, & MacDonald, 1992). In that study, pianists sight-read an unfamiliar piece by Bach while their brain activity was measured using PET. Results supported the idea that brain centers related to motor control, hearing, and vision, all contribute to sight-reading ability. However, musical sight-reading also elicited activation in the parietal lobe – specialized for the conceptualization of spatial relationships – which does not figure into reading of text. This study was also the first of many studies revealing activation in Broca's area, initially thought reserved for speech production (see Figure 4.2).

Recent evidence suggests that brain activity during the playing of music from long-term memory may differ somewhat from sight-reading. In particular, activation in the vicinity of Broca's area was absent during the playing of a Bach concerto from memory (Parsons et al., 2005). Such frontal activations may therefore be important when a person assembles a motor plan 'on the fly' or with little preparation (as is typically the case during speech), but is not necessary for the reproduction of well-learned sequences. In addition, Parsons and colleagues found that much of the prefrontal cortex was deactivated during performance from memory. This finding may seem counter-intuitive, given that the frontal lobe is often associated with challenging mental tasks that incorporate working memory. The authors suggested that such 'higher' functions may actually distract a performer and that when performing from memory one must 'turn off' this kind of conscious thought. Interestingly, a similar deactivation was later found during improvised jazz performances on the piano (Limb & Braun, 2008). Moreover, it is important not to interpret such neural deactivations as evidence that the brain is somehow 'turned off' during a performance. Recent evidence using ERP methods shows rapid fluctuations in brain activity that occur *before* the performer actually makes an error (Maidhof, Rieger, Prinz, & Koelsch, 2009; Ruiz, Jabusch, & Altenmüller, 2009). Thus, not only do performers engage in 'error monitoring,' they are able to predict the outcomes of their actions even before those outcomes occur!

One problem with studying piano performance is that it limits the population one can use. Not everyone can play the piano. However, everyone can sing, though not with equal accuracy. Other research on performance has addressed the neural bases of singing. Perhaps the earliest neuroimaging study of singing was reported by Perry and colleagues (1999), in which they contrasted brain activity during the singing of single pitches with brain activity while listening to pitch sequences. For the most part, singing was associated with activation in brain regions like those used for speech, although singing appeared to involve more right-lateralized activation in the ventral motor region (Figure 4.2 highlights the left motor region). One finding of particular interest was that increased auditory activations were found during singing, as opposed to perception, suggesting that the use of sound for auditory feedback enhances sensitivity to the input. Another study of singing addressed the issue of sequence complexity, and included singing of monotonic as well as more complex sequences (Brown, Martinez, Hodges, Fox, & Parsons, 2004). All tasks involved activations in Broca's area, similar to research discussed earlier. In addition, more complex singing tasks were selectively associated with activation in the planum polare, mentioned in chapter 4.

Because the study of singing includes a vast population, it opens the possibility of exploring to what degree characteristics of music performance exist independent of training. First, the use of motor imagery may not be limited to expert musicians. It seems as though imagining a melody may constitute 'playing' the tune in your brain! In the 1990s, neuroimaging research followed the lead of groundbreaking research in visual imagery to explore whether brain regions involved in imagery matched those used in perception. In general, there is much overlap (see Halpern, 2003 for a review). As might be expected, the auditory cortex (right-lateralized for melodies that are not associated with lyrics) is active for both tasks. Unexpectedly, both tasks also activated the supplementary motor area (just anterior to the motor cortex shown in Figure 4.2), suggesting that people sing or (in the absence of lyrics) hum a tune in their brain when they imagine or hear it. Other recent research suggests that imagined speaking and singing both harness a shared brain area in the vicinity of the planum temporale; this area has been dubbed 'area SPT' denoting the Sylvian-parietal-temporal junction (Hicock, Buchsbaum, Humphries, & Muftuler, 2003). Interestingly, area SPT is left-lateralized, suggesting a music–language link.

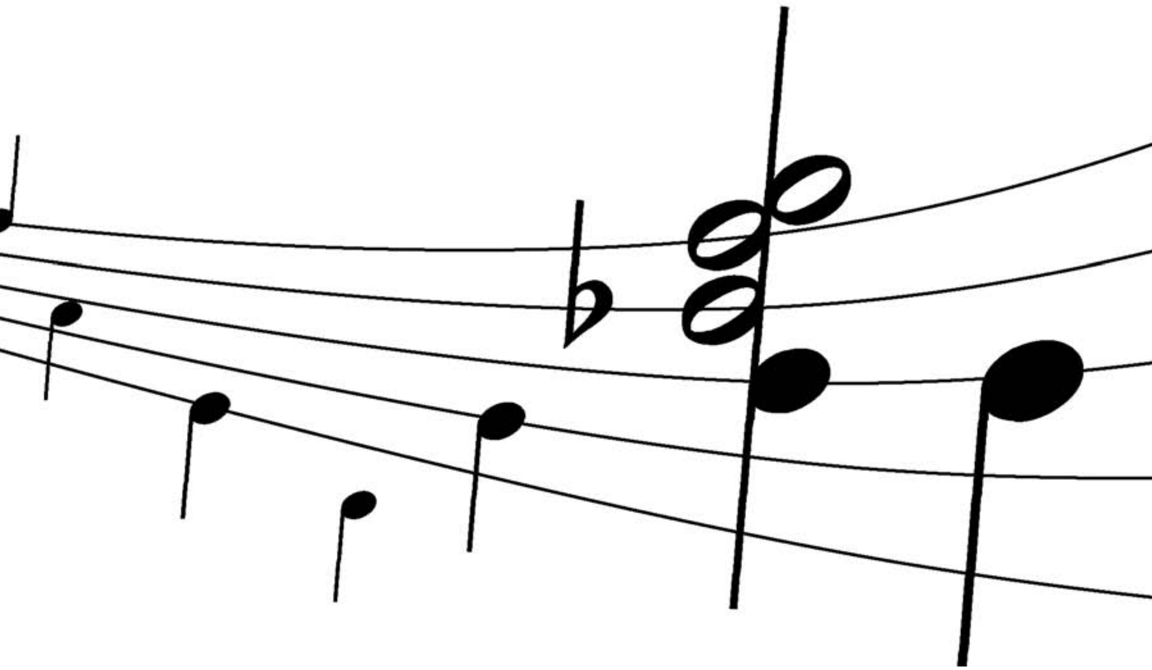
We now turn to the issue, mentioned earlier, of individuals who have difficulty singing in tune. It appears that the brains of 'tone-deaf' people are structured and function differently from the brains of musically 'normal' persons. Structurally, the brains of congenital amusics differ from controls in the distribution of white and gray matter in the pars orbitalis region of the right frontal cortex (see Figure 4.2); those with congenital amusia have proportionally less white matter (Hyde, Zatorre, Griffiths, Lerch, & Peretz, 2006) and more gray matter (Hyde, Lerch, Zatorre, Griffiths, Evans, & Peretz, 2007). White matter comprises neurons that are coated with a myelin sheath, which helps neural signals propagate more rapidly. White matter thus usually serves a kind of 'relay' function for information in the brain; most of the 'thinking' in the brain is served by unmyelinated gray matter. Perhaps significantly, this area is near Broca's area and has been linked to the perception of melodic organization (Levitin & Menon, 2003; see chapter 4). Amusics also respond abnormally to pitch changes, as revealed in a recent ERP study, suggesting that although they may not be fully 'tone deaf' there is nevertheless some problem of 'tone' (Peretz, Battico, & Tervaniemi, 2005). Congenital amusics show greater response to large pitch changes (which amusics can detect) than controls, but smaller responses to small pitch changes (which amusics have trouble detecting).

Coda

The psychology of music performance is a complex and fascinating topic. The performer must coordinate many cognitive, perceptual, and motor functions including planning, execution, the use of feedback, and on-line adjustments of planning. Furthermore, performers expressively deviate from what might be considered a 'typical' performance in order to 'breathe life' into music. Performance is thus a compelling area from both a cognitive and an aesthetic perspective. Musical performance can also be studied from many other viewpoints, and the chapters that follow will examine performance in social context, as well as exploring its philosophical, emotional, and cultural significance.

Part IV

The meaning and significance of music



12 The social psychology of music

The following anecdote was told by Henry Fogel, retired president of the Chicago Symphony Orchestra:

> In 1990, when we were preparing at the Chicago Symphony Orchestra a set of CDs of live performances from the CSO's archives to celebrate the Orchestra's centennial, one performance that we all felt must be included was a Stokowski-led Shostakovich Tenth Symphony. As we listened to the radio tape, there was one lone person applauding after the Scherzo (which was blazingly played). Naturally, for a recording, we edited it out. When the set was released, a few of the old-timers in the orchestra told me that we should have left it in, and explained in the accompanying booklet just who that one person was. It was Stokowski, applauding the orchestra he was conducting!
>
> (Fogel, 2007)[1]

In the opening chapters to this book, we considered sound and music at the acoustic and perceptual levels, but they are more than pure acoustic signals or perceptual events. The sound of clapping can be heard as a mere annoyance interrupting a multi-movement musical work that has not yet been completed; a social gaffe. But on discovering that it is the applause of a world-class conductor who simply cannot restrain himself after a magnificent performance by his own orchestra, the sound takes on a new significance. As performers and listeners, we imbue the certain patterns of sound and silence that we call 'music' with deep meaning; how music takes on the significance that it does is one of its great mysteries. The chapters in the final section of this book address this topic from a variety of perspectives to explore the social, philosophical, emotional, and cultural dimensions of music. The focus of the present chapter is on topics within the social psychology of music.

The beginnings of a social psychological perspective on music are often traced back to Paul R. Farnsworth's 1958 book, *The Social Psychology of Music*, which focused on sociocultural factors influencing musical behavior. The main aim of the book was to explore the cultural determinants of musical behavior, rather than its biological bases which Farnsworth believed had been overemphasized by the 'hereditarians' of the day. The scope of discussion on the social dimensions of music was significantly broadened several decades later in an edited book of the same title by David Hargreaves and Adrian North (1997), and more recently in an edited volume by the same authors entitled *The Social and Applied Psychology of Music* (2008). As it has evolved, this domain of psychology of music has come to more closely focus on aspects of music as they are affected by *social influence*, *social interaction*, and *social perception*. These also serve as the central themes of this chapter.

Hargreaves and North (1997) identified a number of topics that have emerged as the areas of study within this domain. These can be broadly divided into three main strands. There are those topics in which social relationships between performers and each other, performers and their audiences, and between teachers and their pupils are the focus of interest. Then there are those in which aspects of the social world such as gender, social class, age and so on seem to influence such matters as musical taste, choice of instrument, and other preferences. Finally, there are topics which focus on a variety of practical applications of music in such realms as consumer behavior, and the use of music in clinical and therapeutic contexts. In this chapter we will touch on research relevant to one topic within each of these subdomains: the social influence of performers and audiences, the topic of gender and music, and music as a social force with respect to consumer behavior.

Paradigms of social psychology

In the first part of this chapter, we will explore the social relations between people in a variety of musical occasions and episodes. However, before we can review the studies around these themes, it is informative to visit a few influential paradigms in the field of social psychology.

At one extreme there is sociobiology, according to which individual social behavior, that is behavior of individuals in and of groups, is the result of the expression of genetic material through hierarchies of epigenetic rules. At the other extreme lies the social constructionist point of view, owing a great deal to Lev Vygotsky (1978/1930) and more recently to the work of discursive psychologists. According to this view, social psychology is the study of those group processes and the rules and conventions that form part of the knowledge repertoire of individual human beings, so that each person can play an acceptable part in the activities of the group.

Once the dominant paradigm, and now fading away, is the idea that the social behavior of individuals is the result of the establishment of propensities and dispositions to act in certain ways that are activated by stimuli from the immediate environment. This idea, typified by such projects as Berkowitz's experimental studies (Berkowitz, 1967) of what makes people act aggressively, at best offers snapshots cut from a smooth sequence of social actions.

It now seems that the way people act together in the many varieties of episodes that constitute our social lives may be better explained by some combination of the sociobiological and discursive points of view. Instead of trying to explain social episodes in terms of causes and their effects, it may be more illuminating to see what happens as the result of people acting according to their beliefs as to the local rules and conventions appropriate to social occasions of particular kinds. A rock concert and a football match have much in common, but yet the rules and conventions for appropriate behavior are different in many ways – and people know what is appropriate to do in the circumstances. This view will be further explored in this chapter, with application to the behavior of concert audiences.

The power of social influences: Performers and audiences

Social psychology employs scientific methods in order to explore and understand 'how the thoughts, feelings, and behaviors of individuals are influenced by the *actual, imagined, and implied presence of others*' (Allport, 1954, p. 5). In this section, we explore the ways that audiences and performers affect each other, and in the section that follows we shall examine social interaction within various performing groups.

From the audience to the performers

The beginnings of experimental social psychology are sometimes claimed to have been the studies by Norman Triplett (1898) who observed that the presence of other people tends to enhance performance, a phenomenon that later came to be referred to as *social facilitation*. For instance, Triplett found that bicyclists' racing times are faster when racing against others than when racing against a clock. Even children performing simple motor tasks such as winding a fishing reel worked faster alongside a peer than when working alone. Triplett concluded that the presence of other people enhances performance as 'the bodily presence of another contestant participating simultaneously in the race serves to liberate latent energy not ordinarily available' (p. 533).

Robert Zajonc (1965) refined the concept of *social facilitation* by demonstrating that the arousal evoked by the presence of others strengthens the tendency to perform 'dominant' (or highly practiced) responses, leading to performance enhancement in a well-learned or easy task, while at the same time inhibiting performance on a novel or complex task. Indeed, one often anecdotally hears of bands or ensembles that 'feed off' the energy of an audience while sounding rather lackluster during rehearsals. On the other hand, there are fluent rehearsals that are followed by a performance that falls flat. By Zajonc's account, public performances by experienced musicians performing well-worn repertoire should sparkle, whereas the same musicians venturing into an unfamiliar style of music or a newly learned complex piece may give their smoothest performances before the curtains rise!

Nickolas Cottrell and his colleagues proposed that 'the *mere* presence is not a sufficient condition to enhance the emission of dominant responses,' but it is *evaluative* presence that determined the effect of performing tasks in front of an audience (Cottrell, 1972, p. 223). The conditions in which the presence of an audience enhances or interferes with performance include the appraisal by the performer of the audience – and even that of other players in an ensemble. If those present are taken to be friendly and supportive then there is facilitation, especially if the performer's task is well learned. However, if others present are believed to be critical or hostile, the effect on the performance can be dire. This may account for poor performance in front of a somber-looking jury, after having played the same piece 'perfectly' in the practice room for a group of supportive peers.

When the (real or imagined) presence of the audience is detrimental to a performance, the phenomenon is referred to as performance anxiety or 'stage fright.' A large proportion of students and performing musicians (possibly more than half) report being debilitated by anxiety associated with performance (see Steptoe, 2001; Wilson & Roland, 2002 for reviews). Anxiety is not necessarily associated with the size of the audience but is rather brought on by the kind of evaluations that the performer imagines the audience making; in fact the most anxiety-provoking situation is typically the audition, which includes a small but proximal and evaluative audience. Performance anxiety is typically interpreted as occurring because the arousal caused by the audience has surpassed the level at which it promotes good performance and thus begins to debilitate performance, as predicted by the Yerkes-Dodson Law (Yerkes & Dodson, 1908; see Figure 12.1).

It is not surprising then that a popular therapy sought for performance

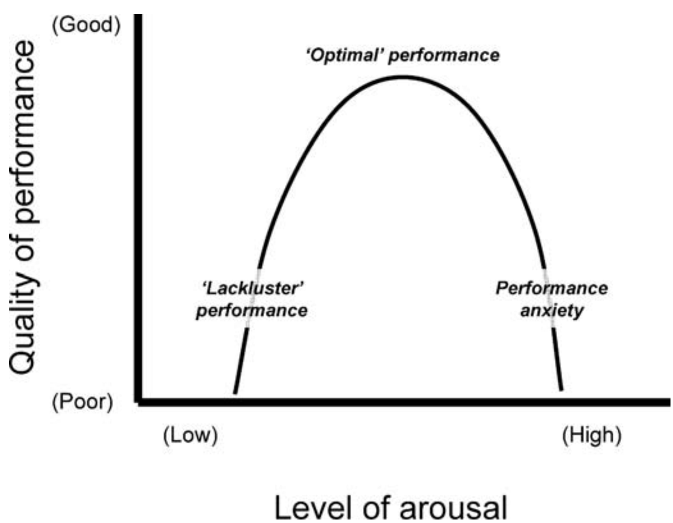


Figure 12.1 The Yerkes-Dodson Law applied to performance anxiety.

anxiety involves administering beta-adrenaline blockers or 'beta blockers.' These treatments artificially lower the performer's arousal level so that it is closer to the optimal level. Of course, such drug treatments have unfortunate side effects, one being that (as shown in Figure 12.1) some degree of arousal is necessary for an 'optimal' performance and so merely reducing arousal may not be the best treatment. Other, possibly more successful, forms of therapy focus on the kinds of thought processes performers engage in (such as cognitive behavioral therapy) so that a more positive perspective modulates the level of arousal a performer experiences. As the saying goes, 'Nerves and butterflies are fine – they're a physical sign that you're mentally ready and eager. You have to get the butterflies to fly in formation, that's the trick!'[2]

Speculating on the psychological basis of performance anxiety, Wilson (1997) suggests that the key factors are fear of humiliation and disgrace that would come from a botched performance. For instance, when adolescents were asked to write about their worst experience in a music performance, Osborne and Kenny found that 60 percent of the total cognitions they coded pertained to 'fear of being negatively evaluated by others and negative self-evaluation' (2008, p. 457). With respect to group performances, Murnighan and Conlon (1991) also found that one of the main differences between successful and less successful music ensembles was the degree of focus on the audience. They observed that 'more successful quartets had a strong internal focus: their primary audience was each other . . . consideration of what an audience desired rarely, if ever, entered into their determination of how to interpret a present composition' (p. 179). However, less successful music groups tended to focus more on audience reactions to their performance, and reported more anxiety prior to performances than other groups.

From the performers to the audience

In a general sense any audience is a crowd, or a group of people who come together in some limited arena of space for a certain length of time, usually with a common purpose. The theory of crowd psychology began with studies by Gabriel Tarde (1843–1904). Following the same agenda, connecting the psychology of crowds with political theory, Gustave Le Bon published his *Psychologie des Foules* in 1895 [1947]. Serge Moscovici (1981) followed the lead of Le Bon in criticizing the Freudian idea that someone caught up in a crowd loses all conscious sense of personal action, and is drawn into a concentration of thoughts and emotions on to a single focus by nothing but suggestion and a kind of contagion. In Moscovici's theory, crowd members are in a battleground of conflicting desires – to have a personal relationship with the person who is the focus of the crowd's interest, and at the same time to be just like that very person. It is not hard to detect these desires vividly expressed in the behavior of the members of the audience at a pop concert.

The etiology of extreme emotions induced during rock concerts has been studied empirically. For instance, interviewing a sample of the 400 people who had fainted at a rock concert, Lempert and Bauer (1995) identified two main ways that the physiological conditions for loss of consciousness are realized in these events. Physical exhaustion and lack of food, coupled with the press of the crowd, induced the kind of oxygen starvation that ended in cases of syncope (fainting) for some people. However, others admitted to hospital reported intense emotions and associated hyperventilation that were most likely simply brought on by the rapid breathing and screaming caused by the exhilaration of being in the presence of the performers that they idolize. Morens (1995, p. 1361) has noted the similarity between the mass fainting at rock concerts and the 'dancing mania' that swept Europe in the Middle Ages. In these contexts, we may see the effects not only of the performer on the crowd but of the crowd on the people within it (and sometimes the performers), as these conditions are seldom brought about when listening to music in a solitary or small group setting.

The fact that performers are seen as well as heard is not a trivial point. The presence of 'live' performers brings much to a performance that cannot be conveyed through an audio recording. In chapter 3, we described a study showing that long-and-graceful versus short-and-swift gestures by a marimba player affected participants' perceptions of the duration of the tones when viewing a video of the performance – even though the tones were acoustically indistinguishable in length (Schutz & Lipscomb, 2007). When participants merely listened to an audio recording of the same tones, there was no perceived difference as a result of the performer making long or short gestures when striking the keys. In another study by Vines, Krumhansl, Wanderley, and Levitin (2006), audience experience of musical tension was predominantly but not wholly determined by what was heard, while experience of phrasing and some structural aspects of the music were influenced by what was seen. As the authors explained, 'musical equivalents of *paralinguistic gestures* (such as head movements, eyebrow raising, and postural adjustments [that are also used in everyday speech]) convey a certain amount of information that *reinforces, anticipates, or augments* the auditory signal' (p. 107, emphases added).

As mentioned in the previous chapter, many studies have shown that performers' movements richly convey information about their intentions, and the expressivity and structure of the music being performed, even when the visual appearance of the performer is abstracted in point-light displays (Davidson, 1993; Schutz & Kubovy, 2009). Although there is not a one-on-one relationship between a specific gesture and a corresponding musical feature, Davidson has shown musicians' expressive movements to be communicative to the audience and other performers, and perhaps also for organizing one's own actions (e.g., to shape phrases). Dahl and Friberg (2007) found that viewers can even identify some emotions (sadness, happiness, anger) solely from the body movements of musicians – in the absence of any sound. The instruments used in the study were marimba, saxophone, and bassoon, and despite the broad differences in the way musicians must interface with percussion versus woodwind instruments, the researchers also found that the performers 'share a common body language' in conveying these emotions musically. Visual information conveyed by live performers is particularly important for musically untrained listeners, who rely heavily on visual cues to interpret performers' intentions (Davidson, 1995). Later in our discussion we will also touch on gestures as a form of communication from performer to performer, and from conductor to orchestra.

In his book on musical acoustics, Hall (2001, p. 401) provides an excellent example of how a musical performance depends on much more than sound alone. When the full orchestra is playing full force at the *fortississimo* level, the sound of the woodwind section is effectively masked by the brass and strings. (*Masking* occurs when a sound is obscured by other sounds of similar frequency, even if as much as 30 to 40 dB softer.) As far as pure acoustics are concerned, the woodwind section could just take a nap in their seats, or pack up their cases and go home! However, the visual appearance of all performers being present on stage and looking busy contributes to the full impact of a climactic passage of music.

The formal audience

Among the alternative approaches to social psychology is the *role-rule* theory of Harré and Secord (1973) that has been used in studying such mass phenomena as football hooliganism. The methodology is based on the principle that social life can be thought of as a series of nested episodes, in which larger episodes such as going to college are made up of smaller episodes, such as the particular courses one takes. In turn each 'taking of a course' consists of a multitude of smaller episodes, such as attending one lecture after another, composing term papers, writing exams, and so on. Each level has its own set of rules and people interact according to certain locally valid systems of rules and conventions. The psychological aspect is brought out by hypotheses about the body of knowledge that must be shared among the people who take part in an episode such as attending an opera performance and the skills with which that knowledge is deployed. This approach and its associated methodology seem particularly apt to facilitate the understanding of the psychology of the audiences at concerts and operas.

Attending a musical event is a complex episode. We have the following components: entering the auditoriums, the dimming of the houselights, the hush, the applause as the conductor makes his or her way to the podium, the performance of the work, the applause, and the encores and the presentation of the bouquets. In each musical culture there are quite definite rules about how the component episodes of the whole event are meant to evolve. Western classical musical performances do not begin with improvisation to which the audience is supposed to murmur approval or disapproval. However, this is a proper part of Indian classical concerts. Though there are definite rules at each level of analysis it is still possible to have individual ways of carrying out the role of the moment. Pianist Lili Kraus used to sing along to her concert performances of Mozart piano concertos. However, this accompaniment could only be heard if seated in very close proximity to the performer (as was one of your authors, RH, at a performance in Auckland).

The people who attend the most formal musical performances may have some of the same underlying psychological desires as the crowd at a rock concert – and some of the same audience members may attend both events – but a different set of social conventions and rules dominate their behavior at each level of analysis. Some of the most striking features of a formal Western classical concert include:

(1) A formal assembly and seating in precisely determined places in the auditorium. There are very precise rules about taking one's seat; for example one does not just sit anywhere at a formal concert. Even in Ancient Greece, the stone seats of the auditoriums had letters incised on them.
(2) Silence falls as the performance is about to begin and remains almost unbroken until a formal sign is given by the conductor that applause is appropriate, usually by a pause and then turning around to face the audience. Clapping, even to convey approval at the end of a particularly stirring movement (as exemplified in the anecdote at the beginning of this chapter), violates the rather rigid conventions for applause. In another example, Henry Fogel (2007) recounts an experience in which a conductor was upset that the audience applauded after the first movement of a performance of Rimsky-Korsakov's *Scheherazade*: 'He turned to the audience, wagged his finger at them, and then held up four fingers and pointed to them, one at a time, to indicate there were four movements!'
(3) The conventions include some of the local variations of how performers and audience accept and express appreciation. Applause is tied in with strict rituals. For example, in Grand Opera, at the end of each of the acts, the leading performers emerge through a gap in the curtain to accept applause singly and then collectively. At the end of the opera something similar occurs, the performers 'taking bows' individually and then as a group. Then the leading female singer goes to the edge of the stage and brings forward the conductor for a special round of applause.

Some of the rules and conventions that make up the body of knowledge of competent concert-goers are set out in the three paragraphs above. There are public reminders of some of the rules, for example the by now ubiquitous request to 'Turn off cell phones, pagers and alarm watches.' 'Stagebill' has published the concert programs for many of the most prestigious performance venues in the United States, and has routinely included their 'golden rules' of concert etiquette in the playbills. Expressed in terms of a kind of contract between audience member and performers, the golden rules remind audience members to 'do unto others as you would have them do unto you.' It goes on to advise against excessive use of perfume and cologne, suggests teaching a child the etiquette of the concert hall, and notes that 'the overture is part of the performance. Please cease talking at that point.' The most forceful rule runs as follows: 'THOU SHALT NOT TALK, or hum, or sing along, or beat time with a body part.' And of course one should unwrap all candies before the concert begins. Psychologically speaking this is a rough sketch of the body of knowledge that concert-goers use to manage their behavior, though rarely are these rules consciously accessed and attended to by audience members.

It should be pointed out that as with many other conventions of concert hall etiquette, undivided attention to the performers is a relatively recent practice in the history of the Western classical concert. This idea was also discussed in our previous chapter on performance. Prior to the early to mid-1900s, concerts were viewed as social events. It was common to talk, visit with friends and admire their fine clothes, and even eat during performances. Reportedly, the pianist Alexander Dreyschock was once reprimanded for playing 'so loud that it made it difficult for the ladies to talk' (Hamilton, 2007, p. 38). Practices regarding audience response have also changed over time. Virtuosi such as Liszt and Rubinstein would have been puzzled as to how their flamboyant performances were being received in the absence of uproarious applause between (and even during) movements. Mahler was one of the first to openly discourage clapping between movements, going so far as to notate in the original publication of *Kindertotenlieder* the following words: 'These five songs are intended as one inseparable unit, and in performance their continuity should not be broken by applause at the close of any of the songs, or by any other interference.' Over time the house lights dimmed over the audience, and the spotlight brightened only the performers' faces – but it has not always been so.

The participating audience

In contrast to the conduct of the formal concert audience, there is the behavior of the participative audience. There is every degree of audience participation in musical events world wide. In an Indian classical concert, the players may be trying out various improvisational forms of the raga that they will develop into a performance. Members of the audience are expected to signify their appreciation of a particularly pleasing innovation, setting the scene for its further development in the performance. Pub musicians encourage a sing-along participation, which, at least in Britain, can be almost formalized in seaside holiday concerts. The words are projected on to a screen and an 'emcee' conducts and encourages the singers. Even in the formal setting of the opera house, it sometimes happens. For example, after the first run through of the Slaves' Chorus by the performers in Verdi's *Nabucco*, the audience is expected to sing it through again. One of the authors (RH) was at a performance of this opera at which the conductor went so far as to rehearse the audience in the preceding intermission. At the 'Last Night of the Proms' in London every summer, Elgar's *Pomp and Circumstance March* No. 1 in its choral version as 'Land of Hope and Glory' is sung by the audience under the direct control of the conductor.

Social interaction in performing groups: Ensembles and orchestras

The coordination of performances by members of large ensembles (in such a way that the audience hears a coherent musical event) is physically complex, considering the time it takes for the sound to reach listeners from different parts of the ensemble. The process is also socially and psychologically complex as the performers must attend to the conductor as well as their own parts, and to the other players or singers. In smaller ensembles such as rock bands, jazz groups and classical ensembles there is no conductor and coordination must be accomplished in some other way. In this section we explore the complex group dynamics of musical ensembles of various structures and sizes.

Ensembles without conductors

A professional musical ensemble (e.g., a string quartet) usually constitutes a self-governing group focused on an intense shared task, which may require working closely together for as long as six hours a day, seven days a week, for many years. Ensembles without conductors face many challenges from within and outside the group.

For instance, Murnighan and Conlon's (1991) study of 20 professional string quartets in England and Scotland identified some main issues inherent to the group dynamics of these ensembles. The classic *string quartet* usually comprises two violins, viola, and cello. The first violin part is typically given the lead and is often the focus of the audience's interest; the second violin mainly supports the melody (for instance, playing in thirds or sixths to the first violin part); the viola usually also plays an inner part but is distinguished by timbre, which gives this part a distinct identity; and the cello usually grounds the musical piece with the bass line. Thus the second violin is not only relegated to a supporting role, but is the only instrument that doubles another instrument and therefore does not have a distinctive musical identity within the group.

Murnighan and Conlon identified three paradoxes that often emerged within these ensembles: the *paradox of leadership versus democracy* (that is, there is often an imbalance in power as the first violinist usually plays a lead role musically and in interfacing with the public, though many musicians join a small ensemble to have an opportunity for more equality); the '*paradox of the second fiddle*' (the second violinist is essential to the success of the group, but is usually relegated to a supportive role in the musical repertoire and audience's perception); and the *paradox of confrontation versus compromise* (conflicts can disrupt the group but are necessary to keep the ensemble unified, for instance, with respect to decisions about the interpretation of a musical work).

It is interesting that parallels can often be found between the musical and social roles that instrumentalists play within their ensemble, contributing to the social dynamics of the group beyond musical contexts. For instance, Murnighan and Conlon (1991) found that in general, first violinists tended to take leadership, not only musically but in administrative roles such as public relations. With respect to the management of these issues or paradoxes, the researchers found that first violinists in the most successful ensembles tended to lead the group while also advocating democracy, while first violinists of the least successful ensembles avoided taking a strong leadership role and were also viewed as less competent leaders by their group. Successful ensembles tended to have second violinists who accepted their supporting status, and were members of groups who expressed their appreciation for the second violinists openly. Finally, successful quartets avoided explicit discussion of the issues outlined earlier but effectively managed them implicitly within their groups. The less successful quartets openly expressed more negative attitudes toward both the group and the music.

Ensembles with different structures may encounter different group dynamics. For instance, Ford and Davidson (2003) studied members of 20 professional and semi-professional wind quintets in the United Kingdom. The technique and timbres of the instruments of the *wind quintet* (which typically consists of a flute, oboe, clarinet, French horn, and bassoon) are more distinct than string quartets, and intergroup dynamics may be quite different from quartets due to the odd number of members in the group. Dynamics inherent in the way the music is scored afford greater opportunities for democracy within the group, such as more equality and interchangeability among musical parts (e.g., the flute is often assigned the melody but sometimes also plays an inner part). There is also more flexibility in seating arrangements for wind quintets unlike the string quartet, which usually adheres to one of two standard seating arrangements. Also, although the flute player was most frequently named as the leader, members of wind quintets were less likely to identify a clear leader than members of string quartets in Murninghan and Conlon's (1991) study. Instead, Ford and Davidson (2003) noted that in most of the quintets, players 'took equal responsibility for leading the ensemble' (p. 62).

As we have seen, the dynamics of a musical ensemble are influenced by a confluence of interacting variables – including those internal to the music (e.g., the particular demands of the repertoire), those inherent in the physical structure of the group (e.g., number of members), and those stemming from influences external to the group (e.g., public perception). There are also the individual attributes of members, such as temperament, leadership skills, proficiency level, and gender (for example see Kemp [1996] on temperaments associated with different instrumentalists). It is interesting to observe some parallels between the musical functions of instruments inherent in the structure of the music and dynamics of the ensemble, and the social roles taken up by the musicians. Are individuals with certain traits and attributes drawn to particular instruments? Or do the instruments that musicians play position performers within a group so that they take on particular roles, and eventually develop the skills appropriate to those roles? Many personality theorists would propose that it is a combination of the two; behavior may depend 'on an interaction between qualities of the person and qualities of the physical and social environment, such that trait-behavior correlations frequently vary with the situation' (Deary & Matthews, 1993, p. 300).

Ensembles with conductors

Expressed in general terms, the task of a conductor and an orchestra, working together, is to create an orderly sequence of musical events according to certain standards of correctness, propriety, and musical aesthetics. The conductor must have the knowledge and skill to create a structured *social entity* out of the group of performers, integrated in various ways with the appreciation by the audience, in order to bring about the musical performance.

Boerner and von Streit (2007) examined the effects of a leadership style referred to as 'transformational leadership,' the three central components of which are charisma, inspirational motivation, and intellectual stimulation. A conductor adopting this style of leadership does not simply enforce his or her own artistic interpretation of a musical work on the group 'but rather conveys it as a vision that the orchestral musicians experience as intellectual stimulation and inspirational motivation' (p. 135), interpreting familiar repertoire with novel and original concepts and a visionary quality. Another important factor leading to good performance is a 'positive group mood' among the orchestral members. Boerner and von Streit's analysis of questionnaires sent to members of 22 symphonic orchestras in Germany showed that these factors do not operate independently, but that the key is in the interplay between the two. Specifically, a transformational leadership style only has a positive effect on the artistic quality of the orchestra if there is also a high positive group mood among the musicians. In turn, a positive group mood among orchestra members only benefits the artistic quality of the orchestra if the conductor strongly embodies a transformational leadership style.

Aside from facilitating social cohesion and inspiring the group, conductors must also coordinate the timing of the performance (as discussed in our next section) and provide artistic direction during a live performance. Music critics often comment on the different 'styles' of conductors. However, studies have shown that conductors' expressive signals – such as the gaze, head movements, and facial expression embodied by Maestra Laura Jackson in

Figure 12.2 Maestra Laura Jackson conducts the Atlanta Symphony Orchestra in an expressive performance of Dvorak's Symphony No. 6 (top) and Prokoviev's 'Classical' Symphony No. 1 (bottom) at the Woodruff Arts Center in May 2008.

Source: Photographs used with permission of Laura Jackson. Copyright © Mark Clague.

Figure 12.2 – are not idiosyncratic but consistent and systematic (Poggi, 2002). For instance, when Maruyama closely compared two conductors conducting the same piece, an analysis of their videotaped performances showed that they used very similar gestures at the same point in the musical work. This implies that 'the character of the musical sound has an intrinsic tendency to arouse a specific type of bodily gesture' (Maruyama, 2007, p. 471). Indeed, Isabella Poggi (2002) proposed that each gaze or facial expression can be mapped onto a precise meaning, and has formulated a lexicon of expressive signals used by conductors. A subsequent study has also shown that vantage point makes a difference as to how well expressive movements are perceived and therefore communicated to musicians. Specifically, a conductor's expressive gestures may be more informative when viewed from the position of the woodwind section (in front of conductor) and first violin section (to the conductor's left) than from where cellists and double bass players sit (to the conductor's right side) (Wollner & Auhagen, 2008).

Timing in an ensemble

One of the most basic requirements of a musical ensemble is to keep time. Though synchronization may seem like a trivial skill to many trained musicians, in fact synchronization of an individual performer with an external rhythm (e.g., a metronome) may constitute a considerable challenge for one who has not practiced the task. Synchronization among many people, all of whom are somewhat variable time-keepers (see chapter 6), compounds this problem. It is of course possible to rely on feedback from other performers to keeping time; however, most ensembles assign one individual with the role of 'time-keeper' with other performers following this person. In rock music the time-keeper is typically the drummer, in jazz it is typically the bassist, and in classical music it is often (but not always) the conductor. The state of synchronization, once achieved, is an all-encompassing phenomenon – a study measuring the EEG response of duetting guitarists suggests that synchronization extends even to the timing of action potentials in the brain (Lindenberger, Li, Gruber, & Müller, 2009).

In the absence of a conductor, players must assume the role of time-keeper. Often one person has a prominent role in the hierarchy of the ensemble and uses specific gestures to direct other performers in keeping time. For instance, in Davidson and Good's (2002) observation of a newly formed string quartet, three types of gestures organized the performance of the group – of which two were related to timing: gestures to mark exits and entrances, gestures to mark dynamics, and 'circular body sway' for timing and expression. Goebl and Palmer (2009) went further to study the kind of timing patterns adopted by performers who take different roles in a piano duet. They found that timing patterns formed by key presses of pianists in the 'leader' position (here, the pianist playing the higher pitched part) tended to anticipate those of the 'follower.' Furthermore, analyses of movements through motion capture (see chapter 11) showed leaders lifting their fingers higher and engaging in more obvious rhythmic head movements – both of which constitute the kind of gestures outline by Davidson and Good.

For ensembles with a conductor (such as in Figure 12.3) it is tempting to assume that the conductor's baton movements 'simplify' the problem of synchronization. Though a conductor may indeed simplify group timing (which is far more challenging for a full orchestra than a string quartet), one should appreciate the subtlety of the conductor's baton. Movements of a conductor's baton are directed toward different points in space that follow a pattern consistent with the meter (see chapter 6 for a discussion of meter). Moreover, the point at which the conductor reaches one of these spatial targets is conveyed by a complex interplay between radius of curvature, movement velocity, and movement acceleration. Geoff Luck (e.g., Luck & Sloboda, 2008) has performed research in which participants synchronize with point-light displays of the conductor's baton; results suggest that no single cue on its own is enough to account for synchronization behavior although acceleration (the rate at which movement speeds up or slows down) may be the most prominent (Luck & Sloboda, 2009).

Whether performing with or without a conductor, the coordination, synchrony, and shared expressivity of performing groups requires what Keller (2001) calls 'prioritized integrative attending.' This idea refers to how a well-coordinated performance requires the careful allocation of attentional resources by ensemble performers – balancing attention to one's own part

Figure 12.3 Maestra Nan Washburn conducts a performance with the Plymouth Symphony Orchestra in April 2008. While orchestral musicians must balance attention between their own parts and the group performance as it unfolds, conductors must allocate their attentional resources among all the performers.

Source: Photograph used with permission of Nan Washburn. Copyright © Catherine Byrd.

while simultaneously tracking the unfolding performance by the group. Both require complex tasks which involve retrieving and activating musical knowledge, executing the motor actions needed to produce the music, monitoring the outcome, and mentally representing the musical episode. Occasionally, the individual performer distributes more attention to his or her own part (for instance when an instrumental or vocal line must be distinct from the rest), and at other times, the performer favors attending to the aggregate (for instance, when blending instrumental or vocal timbre with the group).

The demands of ensemble playing are truly multifaceted and complex, and require the coordination of skills in the cognitive, motor, and social domains. Perhaps it is not surprising to find that conductors are particularly good at prioritized integrative attending. Research using ERPs suggests that the brains of conductors respond to sudden acoustic events that appear in unattended spatial locations (Nager, Kohlmetz, Altenmüller, Rodriguez-Fornells, & Münte, 2003) more so than brains of pianists. In that study participants listened to noise bursts from an array of speakers, with instructions to attend to sounds that came from a specific location. Though not explicitly musical, this experimental setup is directly analogous to a conductor who attends to the first and second violins yet notices a sudden error from the French horn.

Music and gender

Looking at a modern orchestra to make a rough census of the distribution of men and women players, one is still struck by the large number of women in the string and woodwind sections, and the dominant proportion of men among the double basses and the heavy brass, though these proportions have changed a great deal in the last half century. We know very well that there is no discernible difference in the musical competence of people of either sex on any instrument. Gender, the social construction of identity on the basis of the biological distinction between males and females, is continually changing while the biology remains the same. Where does gender differentiation in musical matters begin?

Anthropologists and historians have documented the widespread stereotyping of instruments and performance genres as male-appropriate or female-appropriate. For example, Koskoff (1995) shows that though gender stereotyping in music took different forms in different cultures at different times, the effects are widespread. Occasions for courtship and for ritual have been particularly marked by ideas about gender-appropriate instrument choice as well as type of performance. Funeral rituals among the Ga people of Ghana require certain kinds of songs performed only by women. In Afghanistan typically only women (and children) play a popular instrument called the ‘chang’ (a mouth harp). In a historical study of musical stereotyping in Western Europe, Steblin (1995, p. 144) points to the middle class convention of viewing the virginal and the piano as the most appropriate instruments for 'young ladies' as they could be played to intimate circles of friends and family inside the home. Indeed, prior to the mid-1800s it was considered improper for women to perform in public, and most orchestras would not hire women (O'Neill, 1997).

Gender stereotypes regarding music are already apparent by the time children begin their first music lessons. Music is often conventionally regarded as a more 'feminine' subject and far more girls than boys are engaged in music lessons and activities in the school years. Further, children's choices of instruments are constrained by what they deem to be gender-appropriate. Many studies have quite consistently shown that Western school-age children view flutes, violins, and clarinets as suitable instruments for girls to play, while drums, trumpets, and guitars are regarded as instruments appropriate for boys (e.g., O'Neill & Boulton, 1996, who studied 9- to 11-year-old children in England though similar findings have been shown in younger children).

Some studies have shown that by displaying examples of mismatches between gender and instrument according to the prevailing stereotypes, children's attitudes do seem to change – although the effects are modest and not always in the intended direction. For instance, Harrison and O'Neill (2000) presented live counter-gender-stereotyped models to children, and demonstrated at least a minor shift in expressed preferences among both girls and boys for the conventional assignments of instruments by gender. However, this strategy tended to decrease preference for instruments conventionally regarded as gender-appropriate (for instance, girls indicated less preference for piano after observing a male pianist, and boys indicated less preference for guitar after watching a female guitarist). Another study using video presentations and counter-stereotypical drawings in Australia showed that girls were more willing to consider counter-stereotypical combinations of player and instrument than were boys (Pickering and Repacholi, 2001).

The number of female members in professional ensembles and orchestras is increasing and the scope of instruments represented is wide. However, this may be more true of musical groups in the classical tradition, whereas gender equality has much farther to go in other genres – such as jazz. For instance, in a survey of over 600 students, McKeage (2004) found that far fewer females than males are involved in playing jazz in high school or college. Further, while 62 percent of males who played jazz in high school continued to play in college, only 26 percent of females who played in high school were involved in jazz in college. In particular, the female jazz musicians lacked confidence in their improvisation skills.

Adapting a scale that measures attitudes toward mathematics, Wehr-Flowers (2006) found that females in jazz bands rated themselves as significantly less confident, more anxious, and as having a lower sense of self-efficacy in jazz improvisation than males. Pointing out that most studies have *not* found significant differences in jazz improvisation abilities in males and females, Wehr-Flowers suggests that 'we must then look to alternative possibilities for the gender inequality in the jazz field' (p. 347). For instance, instruments commonly included in jazz ensembles have been stereotypically labeled as 'masculine' instruments, females may not be socialized to feel as comfortable in taking part in jazz rituals such as 'showing off one's chops' as males, and there is an inadequate social framework to support females as the networks through which one obtains informal training in jazz technique and advances one's career are predominantly male (McKeage, 2004).

An area of music in which women are clearly still underrepresented is music composition. Social perception may play a role in the small number of females considered to be eminent composers in many genres of music, as demonstrated in a study employing the 'Goldberg paradigm' (Colley, North, & Hargreaves, 2003). This method was introduced in a classic 1968 study by Goldberg, which showed that the *same* journal articles in various fields of expertise were judged more favorably when attributed to John McKay versus Joan McKay. In Colley et al.'s (2003) study extending the procedure to the musical domain, contemporary music compositions were played to 64 undergraduates who evaluated them on a set of rating scales. Although the results only approached significance, participants tended to give higher ratings on scales related to musical competence when the composers were identified as Klaus Behne and Simon Healy, compared to Helena Behne and Sarah Healy. In another condition, in which a short biography was included (that was identical for all fictitious composers), however, higher ratings were given on some scales for compositions attributed to female composers. As the authors explained, 'where no information other than social category is provided, there is greater pro-male bias' (p. 128). On the other hand, if outstanding biographies are supplied, readers may assume the females to be exceptionally competent to have achieved a high level of accomplishment despite the odds.

In a separate study by the same group of researchers (North, Colley, & Hargreaves, 2003), 153 participants in their late adolescence were asked to evaluate six compositions in the classical, jazz, and new age traditions. In this study, both the names and short biographical passages about the background information and accomplishments of (fictitious) composers were provided in all cases. The results differed somewhat from the Colley et al. study, and the most striking findings of the study were for the jazz excerpts. First, participants clearly perceived jazz composition as a masculine activity – while responses to classical and new age music showed a slight bias in the opposite direction. Secondly, ratings for the jazz compositions revealed clear evidence of 'pro-female bias' by female participants and less dramatic 'anti-female bias' by male participants. Further, the *same* jazz compositions were perceived as more 'gentle' and 'warm' if attributed to a male composer and more 'forceful' when attributed to a female composer, again reflecting preconceptions about male and female composers. As in many other aspects of musicality as reviewed in this section, social factors – rather than raw ability – may account for differences between male and female composers with respect to level of achievement and eminence.

Music as a social force: Consumer behavior

Years ago, the groundbreaking social psychologist Kurt Lewin argued that one's social surroundings play an active role in guiding behavior. Though, as discussed earlier, this view has descended somewhat in recent years, there are strong arguments that music plays this kind of role. These arguments date back at least as far as Plato, who argued that different musical modes may bring about various types of behavior.

Not surprisingly, empirical data show that music can have a considerable impact on behavior, and unlike more explicit forms of persuasion such as verbal messages, it can take place without people being consciously aware that music is guiding their behavior. This point was made in the domain of consumer behavior in a study by North and colleagues (North, Hargreaves, & McKendrick, 1997). As customers milled about a grocery store, music played in the background that encouraged associations with either France (e.g., a soulful accordion piece) or Germany (e.g., brass-laden Bierkeller music). The researchers observed as shoppers walked through an aisle that sold wine. Shoppers' wine preferences shifted toward the country suggested by the music, even though shoppers in general were often not aware of the music at all! The effect of music on behavior can thus be strong and even subconscious.

A number of other studies have shown that music can have a profound effect on the mood and behavior of people in various commercial and industrial settings (see North & Hargreaves, 2005 for a review). For instance, North and Hargreaves (1998) played pop music, classical music, easy listening music, or no music in a university cafeteria. Perceptions of the ambience of the cafeteria varied with music: When pop music was playing the cafeteria was perceived by customers as 'fun' and 'upbeat,' when classical music was playing it was rated as more 'sophisticated' and 'upmarket,' and when easy listening music was playing it was perceived as 'cheap' and 'downmarket.' Further, when asked how much customers would be willing to pay for a list of 14 items sold in the cafeteria, they were willing to pay more when popular music was playing than when no music or easy listening music was playing, and prepared to spend the most money on the same items when classical music was playing. Another study taking place in a university cafeteria showed that music may even affect diners' activity rate. Diners consumed food at 3.23 bites per minute in the absence of music, 3.83 bites per minute when slow music was playing over the speaker system, and 4.4 bites per minute when fast music was playing (Roballey et al., 1985).

Research like the studies reviewed above can give a rather negative impression of music's role as a social force. In this context, we must point out that people seem to use music in this way on their own accord. An ambitious study carried out in the UK (see Sloboda & O'Neill, 2001 for a summary) used the so-called 'experience sampling method' to track uses and functions of music in everyday life. People would be asked to carry with them an electronic pager throughout the day. When the pager sounded (which would occur at unpredictable intervals) participants would be asked to jot down whether music was playing, the kind of music, why it was playing, and the like. In general the results of this experiment support the idea that the typical use of music is similar to the way music is used in a film soundtrack (see chapter 14 for a discussion of the role of music in film). That is, people use music to shape their emotions and behavior throughout the day – a conclusion that has been further confirmed in interviews with people regarding their use of music (e.g., DeNora, 2001). Music comforts us in our sad moments, energizes us when we need to engage in activity, distracts us during mundane tasks, and so on. Music, for most of us, is a social force that we adopt willingly.

Coda

Music is a social phenomenon in several different dimensions. The distinctive character of the music of a culture comes from shared knowledge of the conventions and performance requirements at a particular time and place. However, music is social in another sense. Performances are by people for people. Each performance episode involves the momentary creation of a social structure according to local conventions of propriety and order. Here the process is at least bi-directional, linking performers and audience in various ways, and also performer to performer and audience members with each other. In modern Western classical music, a third element often enters the process – the conductor – actively creating a momentary social order as he or she controls the musicians who in turn interact with the audience not only by the sounds they produce but by visual cues. However, a good deal of ensemble performance successfully unfolds without anyone managing the performance. Sometimes one of the players performs conductorial duties while fully engaged in the performance. Sometimes a kind of collective forms in which features of the performance itself – such as the score, or the relationships between the instrumental parts – serve as a coordinating function. Just as social groups create music, musical performances shape the social groups that they serve to bring into being.

Notes

1 The authors are grateful to Henry Fogel for permission to use quotations for this chapter from his blog *On the Record: Exploring America's Orchestras with Henry Fogel* at http://www.artsjournal.com/ontherecord/
2 This saying has been attributed to Steve Bull.

13 The question of meaning in music

Material things find their places in the world of human beings as much for their significance to the people of a certain culture as for their material attributes. In one sense, a cross is just two pieces of wood. In another sense, it is a religious symbol. It is the same for sequences of sounds. In one sense, they are just sounds like any other environmental noises. In another, they are often the bearers of religious significance, complex emotions, patriotic messages, dramatic story lines, vivid images, and so on. In this and other ways, musical performances have meanings. In this chapter we examine a number of influential theories proposed by philosophers and musicologists as to the nature of these meanings and how mere sounds can have significance for certain listeners. In chapter 14, we will address many of the questions that emerge from this theoretical discussion with empirical studies.

The sciences depend on metaphors as the foundations of theories. Among many other things, the psychology of music must explain not only what kinds of meanings music can have but how that meaning comes about. Musicologists and music critics deploy a vast range of metaphors to try to convey what they take to be the meaning of a musical performance. Psychology of music must make use of these expositions of how meaning is to be understood in the context of musical activities of all kinds.

In a recent review on the role of meaning in music, Ian Cross and Elizabeth Tolbert (2009) highlight two different approaches to understanding meaning in music. One approach is linked to what we will refer to as the *referentialist* perspective. By this view, meaning concerns the way in which a particular symbol system (musical notes, or words in a language) bear reliable associations to things outside the symbol system. This kind of meaning system is referred to as *semantics* in language and some have suggested – controversially – that music may have its own semantics (see Bernstein, 1976 for an extreme perspective on this topic). We will summarize some thinking on how this view might apply to music; though suffice it to say that most are skeptical about how referential music may be. An important implication of the referentialist perspective is that music's meaning is to a considerable degree determined by its context, such as social context (a commercial setting, a concert, a parade, etc., as discussed in the previous chapter). A second way of thinking about music is referred to by Cross and Tolbert as the *aesthestic* perspective. According to this view music exists simply to be beautiful. Though most today would reject such a limited notion of music (much music is designed to counteract standards of 'beauty'), a second aspect of this perspective remains strong: the idea that music's meaning extends no further than the music itself. In other words, music may not be *referential* in the classic sense but may instead be *self-referential*.

We can begin with *program music*, or music that is formally or informally provided with ready-made interpretations and thus is self-consciously referential. Some pieces of music have descriptive titles such as the 'Surprise Symphony' or the 'Leningrad Symphony' or 'Fingal's Cave.' There are story lines that go with the music in operas and oratorios, available moment by moment in the libretto. Popular music is commonly provided with ready-made interpretations in titles and lyrics. During the 1960s and 1970s, 'concept albums' (such as *Tommy* by The Who) unified a set of songs under a common explicit theme or story. These practices suggest that music has the power to refer to something external to itself. A movement or part of a movement might refer to a battle, a storm, a landscape, or some other idea external to the music. Music does seem to have properties which we could call semantic. Using a term drawn from philosophy, we could say music has intentionality (that is, it is *about something*). But is this enough to account for the power music has over the thoughts and feelings of the listener? How might we explain the emotional meaning of a musical work that 'moves' us without explicit references such as these?

Understanding how music can have meaning often draws on metaphors which imply that the power of music is predominantly *emotional*. This can happen in two distinct ways. Music can *cause* or *induce* a listener to actually feel the bodily sensations that are interpreted as distinct emotions such as pride, sadness, joy. Music can also *express* emotions, for instance displaying in sound what a character in an operatic scene is feeling. Musical expression can reach the level of the transcendent, as in great religious compositions, in which feelings of sanctity and awe are the essence of the musical experience. The contrast between the emotion or mood that music induces in listeners, and the emotion a listener understands some musical performance to express, is complicated by the fact that the listener also often gets a certain *satisfaction* from the music whatever emotion it is conveying. Furthermore, such satisfaction occurs even when nothing that could be called an emotional state is felt or understood by the listener. 'Satisfaction' has been claimed to be the most prominent aesthetic aspect of musical experience (Butler, 2004). Some theorists, such as the great musicologist Eduard Hanslick (1891/1986), have defended the idea that it is this satisfaction, an abstract aesthetic experience, that is the only universal aspect of musical meaning.

There is a parallel with the problem of emotion in literature. Words, arranged as narratives, poems, plays, and so on, can both induce and express emotions. Words, properly arranged, can also lead to that peculiar 'satisfaction' that we take in a well-made object. According to Cooke (1959): 'The task facing us [as musicologists and psychologists] is to discover exactly how music functions as a language, to establish the terms of its vocabulary, and to explain how these terms may legitimately be said to express the emotions they appear to' (p. 34). However, most musicologists and psychologists of music would be uneasy, we believe, in taking the parallel between music and language as two major meaning systems too literally. So far we have drawn attention to some similarities, but how do they differ?

Susanne K. Langer (1942) summed up the differences between music and language as systems of meaningful signs as follows:

> The analogy between music and language breaks down if we carry it beyond the mere semantic function in general, which they are supposed to share. Logically, music has not the characteristic properties of language – separable terms with fixed connotations, and syntactical rules for deriving complex connotations without any loss to the constituent elements. Apart from a few onomatopoeic themes that have become conventional . . . music has no literal meaning. (p. 232)

Whatever music is, it is certainly very different in its means of expression from a verbal language. Music does sometimes have some referential meaning, just as some classes of words do. A musical work can stand for something outside the music itself (i.e., something 'extra-musical'), just as a proper name or a description can stand for something outside the language. Nevertheless the way music can have referential import for a listener is not the same as the way that words refer to things.

There are many well-known examples of music imitating the sound of something that is not itself music. In Beethoven's *Pastoral Symphony*, the clarinet produces a sufficiently convincing imitation of the call of a cuckoo to leave no doubt as to which bird song is intended. In Saint-Saëns' *Carnival of the Animals*, the xylophone imitates the sound of clattering bones in *Fossils*. In other cases, nonmusical sounds are interwoven into the music. One of the authors (RH) vividly remembers playing the clarinet in Tchaikovsky's *1812 Overture* when the maroons, large firecrackers meant to represent cannons rather too literally by actually exploding, were set off right behind the woodwind section.

However, when the program notes offer the suggestion that a certain musical object pictures this or that scene or event, it is very easy to conjure up some alternative referent. The storm in Beethoven's *Pastoral Symphony* no doubt has the dynamic structure of a thunderstorm, along with many other natural and social phenomena. It could be rowdy villagers leaving a drunken wedding party, knocking over furniture as they go. The sound of waves breaking on a shingle beach is one possible referent for a certain passage in Debussy's *La Mer*, but not the only one. It could be the wind through a vast field of ripe corn. The titles of these pieces bend our minds towards certain interpretations with far more latitude than our knowledge of language determines the possible referents of words and phrases. Even where lyrics are supplied as an indicator of the emotional message, they may lead one astray if the music and words contradict each other to convey a sort of irony. For instance, in more recent music the British pop group The Smiths often set hyperbolically depressing lyrics to ostensibly happy music (such as in 'Heaven Knows I'm Miserable Now,' set in a major key and at a brisk tempo). Even in vocal music, meaning may not be completely circumscribed by the lyrics.

There are also many examples of composers providing the story in a medium other than words comprising the title or lyrics. In Prokofiev's *Peter and the Wolf*, there is a text that guides the listener in interpreting the themes played by the bassoon, oboe, flute, and other instruments. In Shostakovich's *Leningrad Symphony* we hear the German Army marching ever closer to the city, then in a great battle of national anthems the Russian theme triumphs, and the enemy marches away. The full import of the music requires that the listener has some sense of the national significance of the melodic quotations. Tchaikovsky uses the same device of quoting fragments of the French national anthem 'La Marseillaise' and the Russian Imperial Anthem in his *1812 Overture*. Paul McCartney's 'Mull of Kintyre' is a quotation of another kind, sounding very like a Hebridean lilt, the traditional music of Scotland.

Most scholars in music cognition currently believe that apart from simple association, the emotional force and the narrative meaning of a piece of music works differently from verbal reference, or denotation. Music works by creating some sort of parallelism of structure between the referent and the form of the music, its motion, tension and resolution and so on. In Beethoven's *Sixth Symphony* the music is patterned in the same way as a storm that develops on a humid summer day. We hear a rumble in the distance, then the full fury of the elements and finally the thunder dies away and the sun comes out again. Essentially and strictly in terms of the orchestral activity, we are hearing a crescendo, a passage at double forte, and then a prolonged decrescendo.

Noting that experiments made with vocal music are 'entirely unreliable' as guides in this matter, Langer points out that while 'music is known, indeed, to affect pulse-rate and respiration, to facilitate or disturb concentration, to excite or relax the organism, *while the stimulus lasts* . . . music does not ordinarily influence behavior' (1942, p. 212, emphasis in original). Though people believe that they have the feelings because they have simply been evoked by the music they are hearing, it seems much more likely that they actually have some *idea* of what the music expresses before or during the performance. The phenomenon is as much cognitive as it is auditory and affective. Both composers and performers know how to express emotions, but they do not always actually induce them in the listener or even intend to do so. Nor do they necessarily experience the same emotions that they express; it may hinder the performance if the flautist weeps during a sorrowful passage! Nor is music a conduit between the emotional state of the composer and that of the listener.

Music is a system of symbols, that is meaningful signs – but they are, with some exceptions, nonreferential symbols.

The units of musical meaning

Are the expressive powers of music that are manifested in aesthetic and emotional qualities explicable in terms of the basic relevant sound units that are the elementary bearers of musical significance? Meaning is not likely to reside in individual notes, since a sequence of tones has musical qualities only if the sequences are perceived as a pattern, a *gestalt*, as we demonstrated in chapter 5. Nor is it likely to be found in notes or chords taken out of the contexts of key or harmonic schemes, as we have shown in chapter 7.

One sensible suggestion was made more than half a century ago by Kate Hevner (1936), emphasizing the *temporal* (or time-based) nature of music:

> It is the masses of harmony, the resolutions and progressions, to which meaning attaches, rather than to the triads themselves, and it is the twists and turns of the melody around the keynote, the meaning of the melodic line, rather than certain momentary skips and intervals, which carry the suggestiveness and the meaning of music. (p. 248)

What might be the significance of these resolutions and progressions? Reflecting on what ideas are intrinsic to music, Eduard Hanslick listed 'audible changes in strength, motion, and proportion; and consequently they include our ideas of increasing and diminishing acceleration and deceleration, clever interweavings, simple progressions, and the like' (Hanslick, 1891/1986, p. 10). Music, says Hanslick, can neither signify a particular object of feeling nor the feeling itself:

> It can depict not love but only such motion as can occur in connection with love or any other affect, which however is merely incidental to that affect . . . which moment of these ideas [love, anger, and fear] is it that music knows how to seize so effectively? The answer: motion. . . . Motion is the ingredient which music has in common with emotional states. (p. 11)

How can Hevner's and Hanslick's observations be tied in with the evident meaningfulness of many if not most musical performances? How could formal structure and motion be the sources of emotional understandings and even emotional experiences? How could these formal properties underlie the way many musical events are supposed to depict places and happenings? It is to these questions that we turn our attention in the next section. Having considered some of the ways in which musical and linguistic meaning are similar and different, we now turn to some theories of meaning as proposed by several philosophers.

Theoretical accounts of meaning in representational and nonrepresentational music

Theoreticians concerned with the topic of meaning have formulated theories of meaning in music as well as for many other kinds of representations (see Harré, 1997). We will examine three accounts of what meaning might be in the context of music.

J. Hospers (1964, pp. 44–49) distinguishes between 'pure' music and 'representational' or program music, a distinction pioneered by Hanslick (1891/1986). According to Hospers, the difference is one of degree. *Program music* is 'defined as music which evokes in most listeners the impression of specific objects or situations' (p. 44). It is partly a matter of the composer's or the improviser's intentions, since Hospers goes on to add that representational or program music is 'designed to invoke' some specific object or event. *Pure music*, on the other hand, does not evoke any assigned specific object for the listener. However, Hospers does not offer any theory of how representational music is able to fulfill the composer's intentions. As suggested before, perhaps it is brought about by the title or program that accompanies the music. Mendelssohn called his piece a '*Hebridean Overture*,' delimiting the hearable references to the scenes encountered on a sea voyage. Debussy's lyrical titles *Reflets Dans L'Eau* (Reflections in the Water), *Clair de Lune* (Moonlight), and *Jardins Sous La Pluie* (Gardens in the Rain) all impress on the performer and listener a particular set of vivid images.

J. O. Urmson (1973) takes the distinction between representational and nonrepresentational music a step further. He points out that the use of a musical phrase to refer to something outside the music (the clarinet's imitation of the cuckoo in Beethoven's *Sixth Symphony* for example) is asymmetrical. Within the symphony we can say the motif represents a cuckoo, but a springtime cuckoo call in the Vienna Woods does not represent anything in the symphony, even to Beethoven. Thus this type of representation does not work in both directions. Urmson elaborates on the idea of composer's intention in a formal definition:

> 'A represents B' means 'A is auditorily similar to B and X intends A to represent B.'

This definition introduces the musical context indirectly. The composer has created the passage to provide a musical footing for the representational phrase or note. Again, it must surely be the program notes or the title or something else that disambiguates the possible objects represented by the music. Urmson makes this point in arguing that 'extra-musical evidence [i.e., extrinsic to the music] is normally necessary' (1973, p. 137). Familiarity with a particular musical tradition also plays a part. Beethoven, Rossini, and Mahler all 'do' storms in a recognizable way. Knowledge about a particular composer may also play a role. A symphony by Mozart in the key of G minor, for instance, may take on greater significance for the listener who knows that Mozart had an affinity for major keys and rarely wrote in minor keys; he wrote only two symphonies in minor keys (No. 25 and 40) and chose G minor for both.

Hospers and Urmson take for granted that knowledge of the world of nonmusical sounds, situations, events, and so on shapes the way the music can be interpreted. On this view, the aesthetic and emotional appeal of non-representational music remains a mystery. Roger Scruton, on the other hand, reverses this direction of influence. According to Scruton music reveals to us exactly what we have put into it from our own resources. These resources come from our experience of ourselves. In Scruton's (1983) words: 'It is *our experience of ourselves*, rather than a scientific representation of the world, which both prompts and explains the metaphors we apply to music' (p. 86, emphasis added). Thus our experience of movement, rhythm, and harmony should be explicable by reference to a projection of something subjective on to an auditory object, for example, the ticking of the clock as we hear it, rather than the movement of the gears and escapement of the machine itself. In this way our personal experience makes possible a particular interpretation of a heard sound pattern.

On 'movement,' a key component of musical experience, Scruton (1983) remarks 'it is ourselves as agents [that is, the sense of ourselves as originating sequences of events] – rather than any purely geometrical idea of space – which underlies our experience of musical movement' (p. 86). He explains the perception of rhythmic patterns in a similar way in the following statements: 'The perception of rhythm involves imaginative transfer of the kind involved in metaphor' (p. 90). Thus, '[w]hen we hear a rhythm we hear sounds joining to and diverging from one another . . . in a manner familiar from our knowledge of human movement' (p. 90). Rather than the music driving the dance, the dance drives the music! Before the minuets of Haydn and Mozart and the waltzes of Strauss, there were people dancing to threesome rhythms. A great deal of current popular music is performed as part of a total experience that includes rhythmic movement as an essential core.

The roots of our experience of harmony are different. 'Harmony' says Scruton (1983) 'belongs not to the material world of sound but to the intentional [meaningful] world of musical experience' (p. 93). Harmony appears in the 'tension, transition and resolution' to surrounding chords. The matching patterns (or isomorphisms) between the structure of emotional experience and the structured movement of a musical object is a synthesis. It occurs because we search in the musical object for something that has its origin in ourselves.

Although it is generally thought that the capacity of music to refer to extramusical events may be fairly limited – even within program music – there is some recent neuroscientific evidence that supports the notion that certain general semantic associations may be triggered by musical sequences (Koelsch et al., 2004). In psychological terms, a musical sequence may prime one's memory for a particular kind of word meaning (see chapter 5 for further discussion of priming). As Koelsch and colleagues put it, 'it seems plausible that certain passages of Beethoven's symphonies prime the word *hero* rather than the word *flea*' (p. 302, emphasis in original). They put the idea to the test in an experiment in which they measured neural responses using event-related potentials (see chapter 4). The paradigm they adopted focused on 'semantic incongruities.' In language, this involves ending a sentence with a word for which the meaning conflicts with the overall context (also discussed in chapter 4). In a pilot study, the authors identified musical sequences that were associated with certain words. They then presented sequences to a different set of listeners in which a musical or linguistic sequence (for the music, for instance excerpts by Strauss and Schoenberg) would be followed by a target word (such as 'narrow'). The authors found similar neural responses when the target word did not match its context, regardless of whether the preceding context was a melody or a sentence.

None of the accounts given above is entirely adequate when considered alone. Though too schematic to count as substantial theories, they do point the way to deeper theorizing. In chapter 14, we will survey the experimental work on how music can have meaning sufficient to invoke or express definite emotions and moods. But first we must complete our preliminary foray into this difficult topic by drawing some important conceptual distinctions needed to understand current empirical research.

Cognitivist and emotivist theories of the emotional powers of music

As we pointed out at the beginning of this chapter, music and the emotions are entangled in two very different ways. A musical performance can be the cause or the occasion for the listener and sometimes the performer to actually experience a certain emotion. Patriotic music might induce a bodily feeling easily interpreted as national pride. A melancholy love song might literally 'bring tears to one's eyes.' However, music can also be used to express an emotion that is understood by but not induced in the listener or the performer. A powerful aria by the Queen of the Night expresses jealous rage in Mozart's *Magic Flute* but we can be fairly sure that no one in the audience experiences jealous rage at that point in the unfolding of the story on the stage. One does not need to be Welsh, to have no difficulty in understanding what is expressed by the singing of 'Mae Han Wlad Fy Nhadau' ('Land of our Fathers') as the Welsh rugby team takes the field, though there may be no swelling of tear ducts or racing of the pulse except perhaps for those who are Welsh. This feature of musical expression is particularly striking in opera, where the unfolding of the story requires the listener to understand the emotion overcoming a character, without necessarily sharing it. When The Beatles brood on 'chagrin d'amour,' the loss of love, in 'Yesterday,' we can sympathize without empathizing.

Music does not generally induce emotions by causing people to feel the emotions it expresses. Of course, there are individual reactions, specific occasions for effects that are sometimes no more than simple conditioning. In these cases it would be acceptable to think of the relation between hearing the music and experiencing the emotion as a case of cause and effect. A person might have acquired a habitual emotional reaction to a certain tune by the regular association with an emotion or mood, such as the patriotic pride that is evoked by a National Anthem. The idea or image of a funeral is associated with a certain tempo, the ponderous beat of a funeral march. This association enriches the listener's experience when the double bass introduces the third movement of Mahler's *First Symphony* in the manner of a funeral march, though the melody turns out to be *Frère Jacques*! But the music might have been associated on just one striking occasion with an image, a feeling or scene, so that hearing the music *brings to mind* the image, the feeling, or the scene once again. It is akin to what J. B. Davies (1978, pp. 68–70) referred to as the 'Darling They're Playing Our Tune' phenomenon, an emotional response to music through association with a specific event.

We follow Kivy (1990) in referring to the view that music *induces* emotions in listeners as the *emotivist* position, and to the view that music *expresses or represents* emotions which listeners comprehend as the *cognitivist* position. In the psychological literature these positions have often been presented as rivals as if we must choose one or the other. However, it is likely that music is such that we need both positions to do justice to the relation between musical experience and emotion.

The distinction between cognitivist and emotivist positions in the psychology of musical experience is complicated by the fact that there are two contrasts in play. Cognitivists say that we understand a certain musical performance to be expressing an emotional state, by a kind of 'description,' for example a matching (or isomorphism) between the structure of the music and the structure of an emotion. Emotivists say that the musical object is causing an emotional state in the listener through physiological responses. However, the emotivist theory has been proposed in two versions, related to distinctive theories in the general psychology of the emotions. Reductionists say that an emotion just is the bodily state that is caused by the musical performance, while nonreductionists hold to a semi-cognitivist view, that only as these induced physiological states are interpreted according to certain schemata are they experienced as emotions (or moods). The same distinction appears in discussions of the *means* by which music achieves the expression of emotions.

Aside from some special cases such as the conditioned responses discussed earlier, there seems little reason to dissent from the idea that emotions are the result of *interpretations* of induced physiological states. That is, the explanation of emotional experiences in relation to music is inherently cognitive. Some research in psychology favors the cognitivist view, but in such a way that finds a place for induced physiological states. In brief, it seems to be the case that specific emotions are experienced through the interpretation of physiological states according to a fairly complex cluster of cognitive schemata, including the meanings of emotions as displays of social acts. For example, a display of anger can be read by a transgressor who sparked it off as a protest at some kind of offense. According to this view the interpretative aspects of emotions are learned as local conventions, and likely to be culturally specific.

In sum, the psychologist of music is confronted with two contrasts. Are emotions essentially physiological or are they essentially cognitive? Is the emotionality of music its power to induce emotions in listeners or is it its power to express emotions, though not in the way that poems and stories usually do? The truth of the matter seems to lie in some judicious blending of these contrasts.

Theories of how music and emotion are related

Having addressed some theories of meaning that encompass many contexts for meaning and introduced the cognitivist–emotivist debate, we now turn in our final section to theories that more directly address the problem of how to account for the relationship between music and emotion. We draw mainly from Susanne Langer's (1942) *Philosophy in a New Key* and musicologist Leonard Meyer's (1956) *Emotion and Meaning in Music*, while also touching on philosopher Peter Kivy's and musicologist David Huron's ideas more briefly. Most research in the area of psychology of music, even in the present time, is grounded in the seminal works of Langer and Meyer. Indeed, their work anticipated the recent shift in psychological research on emotion, from accounts dominated by the physiology of bodily reactions, towards accounts that include semantic and social aspects of emotional experience and displays.

Peter Kivy's strong cognitivism

In his spirited defense of a cognitivist position against emotivism, Peter Kivy (1990) argues that not only is it not true that music induces emotions in listeners, but that it could not do so. The way sound is heard as music precludes this possibility. His theory rests on the principle that understanding anything must be relative to perceiving it 'under a description.' Whether a figure is seen as climbing a slope or slipping back down it will depend on the story line that accompanies the perception. If the story line is 'intrepid mountaineer' we will see the person trudging upward. But if the story line is 'initiating an avalanche' we will see the figure as going down to disaster! According to Kivy (1990, p. 99) 'evidence of understanding music is the ability to describe, and enjoying music is a direct function of understanding it.' He develops this thought with the claim that the level of enjoyment of some musical object is the level of available description. A description displays that of which we are aware, that is the musicality of the piece for this listener. For example, if the listener describes the piece in terms of the subtlety of the variations invented by the composer, the music means something different from the understanding of someone who simply notes that it is in a minor key. However, it surely cannot be the case that every activated cognitive schema must be voiceable. Kivy (1990, pp. 127–128) notes this objection without resolving it.

Kivy's theory seems to hint at a much deeper theory of how music can have intentionality. To this end, both Meyer (1956) and Langer (1942) have important insights to offer psychologists.

Leonard B. Meyer's theory of emotional meaning of music

We now turn to musicologist Leonard B. Meyer's well-known account of the source of emotion experience in listening to music. From the outset, Meyer rejected any attempt to reduce emotional experience to physiological states and processes (Meyer, 1956, p. 13). Such states do not differentiate affective from nonaffective states. The difference must be sought in mental activity, in the realm of meanings. Nor is affective musical meaning to be looked for in extramusical experience (i.e., with reference to objects and events outside of the music). Thus Meyer's account, unlike some of the other views discussed earlier, does not focus on musical meaning derived from some relation to a source external to the music itself – such as a storm, lily pads on a pond, or the pain of unrequited love – but explores musical meaning in a sense that is more intrinsic to music.

Meyer's account of meaning in music is derived from the principles of Gestalt psychology, particularly the Principle of Good Continuation. He defines 'musical meaning' as follows: 'one musical event (be it a tone, a phrase, or a whole section) has meaning because it points to and makes us expect another musical event' (Meyer, 1956, p. 35). The technical term for this property of event (or thing) is 'intentionality.' However, he notes that whether a musical event has intentionality for a listener depends on that listener's knowledge of musical style.

Meyer (1956, p. 37) suggests a pattern of development of the apprehension of musical meaning as the listener follows the evolution of the music. *Hypothetical* meanings arise during an act of expectation, if this [i.e., the antecedent] has happened that [i.e., the consequent] is to be expected. *Evident* meanings are attributed retrospectively to the antecedent after the consequent has occurred and the relation between antecedent and consequent has been perceived. *Determinate* meanings arise out of the relationship between hypothetical and evident meanings as the music undergoes further development. In other words, we do not know what the significance of a tone or phrase is until we have heard what follows it in the musical process.

Essentially Meyer's account rests on a general thesis that 'emotion is evoked when a tendency to respond is inhibited [resulting in a *tension*]' (Meyer, 1956, p. 22). However, this broad hypothesis about the sources of emotion in general needs qualifying to differentiate it to deal with the special case of music. Given Meyer's assumption that musical experience is nonreferential (i.e., its meaning does not rely on its capacity to refer to something outside itself), the source of expectations must lie in the nature of the musical performance itself, the ordered sequence of heard tones. Wherever there are the beginnings of a gestalt pattern that are not yet completed, there must be tension. Musically this is easily identified in various forms of incompleteness, such as the tantalizing holding back from resolution to the expected tonic in classical Western music.

To Meyer, then, tensions arise from expectations aroused by patterns that in some way are experienced consciously or unconsciously as unfulfilled or incomplete, and which may be fulfilled or thwarted. This can only happen against a background of expectation of continuity. Meyer lists a number of such continuities, the violation of which arouses emotion: melodic continuity, rhythmic continuity, and the basic and universal diatonic scale. Violations of the principle of diachronicity include chromaticism and the minor mode.

In previous chapters, we have given an account of the cognitive processes that can be used to explain the phenomena of expectation and its frustration or fulfillment that Meyer describes. The Gestalt theorists, whose ideas we discussed in chapter 5 and on whose work Meyer draws, have given some strong working hypotheses of how this could happen. In turn, Meyer's views have inspired many subsequent theories. Narmour's (1990) Implication-Realization model complements and expands on Meyer's views, and is summarized in chapter 7. Another theory that expands on Meyer's work is that of Huron (2006).

Musicologist David Huron has proposed a theory of expectation referred to as *ITPRA* (Huron, 2006). The acronym refers to successive stages by which we anticipate events and then respond to the manifested outcomes of observed events. The stages comprise *imagination* (what will come?), *tension* (waiting for it to come), *prediction* (did it come . . . or something else?), *reaction* (how do I feel about what came?), and *appraisal* (what are the implications of this outcome?). Though his is a theory of musical expectations, Huron grounds his views in evolutionary theory and as such, it is a theory that could respond to expectations for any temporal sequence – not just music. ITPRA is similar to the theory of Meyer in that both rest on the assertion that musical experiences derive to a great degree from the way in which our expectations are met or thwarted. But striking differences lie in the bases for these expectations. Meyer's theory (and even more so Narmour's theory) are fundamentally *deterministic*, in that they propose basic musical forms (or archetypes) that lead fundamentally to a certain kind of expectation. By contrast Huron focuses on the kinds of *statistical* regularities in music. In ITPRA, expectations are formed based on the kind of experience we have and are inherently more flexible. As an example, Meyer proposes that the tendency for a large pitch leap (e.g., an ascending octave) to be followed by a smaller interval in the opposite direction (e.g., a descending major second) is an inherent property of music. Huron, by contrast, proposes that such forms are a byproduct of a statistical property known as *regression to the mean* (von Hippel & Huron, 2000). Small intervals are more common than large intervals, and instruments tend to stay close to an average pitch level. Thus when a melodic line makes a large movement away from this average pitch, the following interval is more likely (based only on probability) to be both smaller and to go in the direction of the average.

Susanne K. Langer: Isomorphisms of structure

In everyday life tensions are frequently unresolved, while in music *resolution* of tensions is at the very heart of music as an experience. In ordinary life, the source of tension and the mode of resolution may be very different in general character (a quarrel may be resolved by a brisk walk in the garden). In music, however, tension and resolution make use of the very same means, the developing pattern of an ordered sequence of tones. Meyer notes that this thesis leaves open the question as to *how* musical events 'arouse and inhibit tendencies and thereby give rise to emotions' (Meyer, 1956, p. 24). It is here that Susanne K. Langer's thesis of structural isomorphism between the forms of musical objects and those of emotional and other processes comes in to complement Meyer's theory.

Philosopher Susanne Langer's (1942) theory of how music can have meaning was one of the sources of the Meyer/Narmour Expectation/Realization proposal as the basic mechanism of musical experience. Her theory bases the listener's understanding of music on isomorphisms. In Langer's sense, *isomorphism* refers to patterns of similarity between the structure of the musical rhythms and melodies and the structure of other cognitive, discursive, and above all emotional 'objects' to which the music reaches out. In so far as a piece of music shares a common form with a story, a scene, a storm, a battle, or a more intimate emotional episode, so the music is capable of expressing it.

This neo-Schenkerian theory bridges the gap between the formal analyses of Schenker and Lerdahl and Jackendoff set out in chapter 7, and the phenomenology of musical experience, in explaining how music can have meaning – particularly emotional meaning. Langer's theory is further developed in her study of the general features of symbolic forms, the characteristics that endow all sorts of material things with various kinds of meanings. The structure of paintings, sculptures, landscapes, even cars can be isomorphic to emotions, story lines, and so on. Think how the shape of a Jaguar car is isomorphic on the one hand to the shape of its eponymous animal and on the other to a sense of rapid progress of both through space (Langer, 1953).

In short: Langer's answer to how music, as a nonreferential symbol or system of symbols, can express anything, is that the form of a musical event can be isomorphic with the form of something nonmusical. It might be a story, a dance step, or, in the case being discussed in this chapter, an emotional experience.

Langer draws authority from several authors who seem to have anticipated this idea. For example, she alludes to Wagner's remark that music does not express this or that person's emotion but rather what the emotion is itself. Music, says Langer (1942), is the 'formulation and representation of emotions, moods, mental tensions and resolutions – a "logical picture" of sentient, responsive life, a source of insight not a plea for sympathy' (p. 222). Music is not a language in which emotions and life events might be described, but 'expresses primarily the composer's knowledge of human feeling' (p. 221). Langer puts this in terms of a contrast between self-expression and logical expression. The only plausible theory must be based on the idea of logical expression, that is music as significant form: '[M]usical structures logically resemble certain dynamic patterns of human experience . . . *what music can reflect is only the morphology of feeling*' (pp. 227, 238, original emphasis). Quoting from Carroll Pratt's (1931) *The Meaning of Music*, Langer writes:

> Music has its special, purely auditory characters, that 'intrinsically contain certain properties which, because of their close resemblance to certain characteristics in the subjective realm, are frequently confused with emotions proper'. But '. . . these auditory characters are not emotions at all. They merely *sound* the way moods *feel* . . .'
>
> (Quoted in Langer, 1942, pp. 244–245, original emphases)

To complete her account, Langer introduces the concept of the 'unconsummated symbol.' In music 'we have an unconsummated symbol, a significant form without conventional significance' (1942, p. 241). It can be completed in all sorts of ways. 'The power of music,' she declares, 'lies in the fact that it can be "true" to the life of feeling in a way that language cannot; for its significant forms have that *ambivalence* of content which words cannot have' (p. 243, emphasis in original).

We can represent Langer's theory of music as a flow of unconsummated symbols diagrammatically, as shown in Figure 13.1. In order to express something musical or nonmusical, the structure heard in music must match the structure of what is expressed.

In this way, music can give expression to something that is not music. To establish this claim, detailed research would be needed to reveal the relevant structures of all sorts of 'external,' nonmusical objects or events. An emotional episode can have a rising tension and some kind of resolution, whether violent or calm; a storm can be anticipated by a rumble of thunder, then the

Music				Emotion, story etc
↕				↕
Structure 1	↔	isomorphism	↔	Structure 2

Figure 13.1 The mutual structural mapping of music and emotion.

flashes of lightning and the roar of the rain and finally the thunder dies away, and all we hear is the tinkle of rain drops falling from the leaves into the pools beneath. Love affairs, duels, hunts, and so on are all structured broadly in this way.

Summing up, we can say that according to Meyer (1956) fulfillment and frustration of expectations induces tension against a pre-established background of 'normal' development of the music we are hearing. Following Langer (1942), music can express emotions because there are structural isomorphisms between features of music and features of emotion episodes. For example, tension/relaxation/delay in the musical form matches tensions/relaxation/delay in the mental state, story line or whatever it might be that the music expresses. Given this general theory, our task now is to use empirical studies to identify some of the ways that the structure of music can be relevantly isomorphic to the structure of an emotion, a mood, a legend, or an everyday story. This sort of research should lead to a variety of structural theories, each appropriate to its particular domain. We turn to a survey of empirical work on this topic in the next chapter.

Coda

Theoretical accounts of the intentionality of music, in particular of the emotional power of music, take us in two contrary directions. Some pieces of music, and even certain musical phrases, become *associated* with emotions in the course of life, needing no more than the psychology of conditioning to explain it. This simple link can extend from the unique experiences of individuals to the emotionology of a whole culture. However, according to several theorists whose ideas were explored in this chapter, there is a deeper dimension: *structure*. This is where the fundamental psychological power of music lies.

Though much has been learned about the kinds of meanings that music can have for listeners, including expressing and sometimes inducing emotions, many questions remain to be answered. Music is great art. Its appeal is to that part of the psyche to which any form of great art appeals. But how does this happen? What processes occur in the listener that sustain the possibility for the induction and expression of emotion? The emotional content of musical experience ranges from the most banal of personal associations to the most refined aesthetic and even mystical evocations of the very nature of the cosmos. In the next chapter we turn to some experimental studies on the topic of emotion in music to explore these questions.

14 The emotional power of music

'This song arouses so strong feelings in me that I cannot listen to it more than once a year.' '. . .It made me ecstatic, inconceivably exhilarated, everything concentrated to a single now'. These are two responses given by participants in Gabrielsson and Lindstrom's studies, when asked to describe 'the strongest, most intense experience of music' they had ever had as a listener or performer (Gabrielsson, 2001). How and why does music rouse such strong emotions in us? What gives music its affective power?

In the preceding chapter we worked our way through various competing and complementary accounts of musical meaning, particularly on meaning that is emotional. Our fundamental working distinction between music as a medium for *expressing* emotions and music as a stimulus for *inducing* emotional states reappears in this chapter. We now turn to some of the classic and current experimental research related to these theoretical developments. In their compilation of theory and research on music and emotion, Juslin and Sloboda (2001, p. 3) observe that the topic of emotional aspects of music 'has been seriously neglected during the last decades' and regard the work in this area as being in a 'pre-paradigmatic' phase. Nonetheless, psychological studies have made some important contributions towards resolving some of the issues surrounding this intriguing topic.

It is important to consider that the psychological definition of emotion is more nuanced that one might expect. Colloquially, the terms 'emotion,' 'feeling,' and 'mood' are often used interchangeably, but the standard definitions within psychology (though overlapping) are not interchangeable. In particular, emotions are thought to include feelings (a subjective experience, e.g., sorrow), but also to involve appraisals of a situation (e.g., one's cat just died), to be focused on an object (the cat), and to be associated with physical expressions of the emotion (crying). It may well be the case that the 'emotions' associated with music are in fact emotions in the broader psychological sense (e.g., Juslin & Västfjäll, 2008) though an alternative proposal states that music elicits strong feelings but only generates emotions through extramusical associations (e.g., Konečni, 2008).

In the psychology of emotion as a leading aspect of musical experience, we can recognize two major research questions that might be answered by the results of empirical research. The first concerns the question of discerning how musical emotions are communicated – this can be thought of as the *expression of emotion*. A second issue involves the kind of emotional response that music may induce in the listener, the *induction of emotion*. This second issue invokes the controversy mentioned in the previous paragraph. Certain studies we will discuss examine whether listeners can recognize what emotion a piece of music is intended to communicate. However, this recognition is not the same thing as actually feeling an emotion. In order to understand whether music actually triggers 'felt' emotions we will review research that includes measures of physiological responses while people are listening to music to see whether there are any characteristic neural responses to music reported to have distinctive emotional qualities (Harré & Parrott, 1996). We may also use the existing findings on the physiological responses of the body that are associated with the experience of emotions to attempt to map one onto the other.

There is also the question of the practical implications of music's capacity to express or induce emotion. At the end of the chapter, we turn to a real-world context in which these effects are particularly powerful: the role of music in film.

Expression of emotion in music

One way of examining whether formal attributes of music lead listeners to specific emotional experiences is to look for patterns of reliable relationships between features of music and the emotions they express or convey as these are reported by listeners (Gabrielsson & Lindstrom, 2001). In so doing, we can address such questions as: Do listeners generally agree on the emotions expressed by a piece of music, or are their interpretations highly variable and idiosyncratic? What specific features of music convey or express emotion? An early and important study along these lines was carried out by psychologist Kate Hevner (1936), whose ideas were presented in the previous chapter. The methodology she employed and the results she arrived at are of lasting significance, so we will look more closely at her work here.

General musical features

Two experiments summarized by Hevner (1936) focused on four features of music: These were major and minor key, rising and falling melodic line, firm or flowing motion in rhythm, and simplicity and complexity of harmony. For each feature, two recordings of the *same* piece of music were created (e.g., the same composition played in 'firm' or 'flowing' rhythm). In the case of rising-falling melody line and simple-complex harmony, alterations had to be made to the musical works to create the two versions. Participants were presented with only one version for each pair, and were asked what the music expressed to them. They indicated their responses by selecting adjectives from a list of

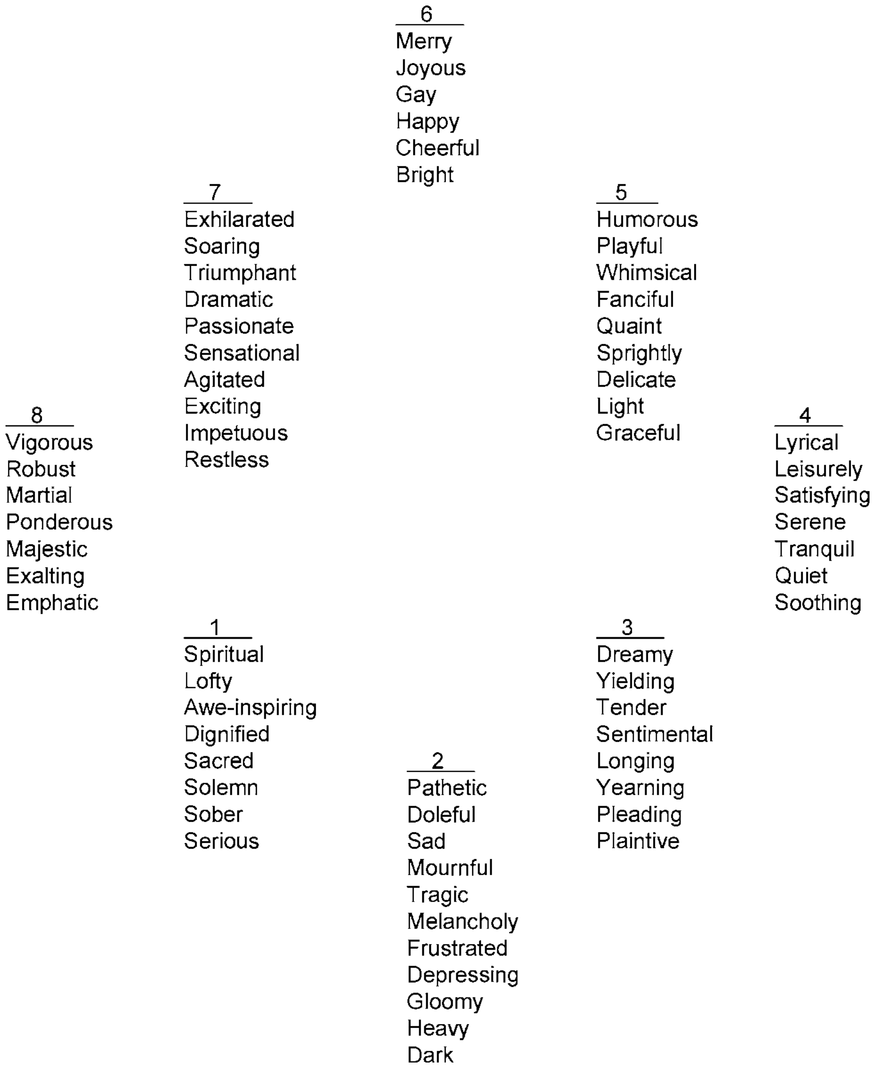


Figure 14.1 Semantic scheme used by Hevner (1936, p. 249).

Source: From *American Journal of Psychology*. Copyright 1936 by the Board of Trustees of the University of Illinois. Reproduced with permission of the University of Illinois Press.

66 adjectives assembled by Hevner (see Figure 14.1). The focus was on the music rather than any subjective state that the participant might have fallen into. Thus Hevner's research concerned the *expression* of emotion musically but not the induction of emotion by music.

There was general agreement as to the affective tone expressed by the musical features, among both musically trained and untrained listeners. The main findings that emerged from the study can be summarized as follows (Hevner, 1936, p. 268):

(1) The major mode is 'happy, merry, graceful, and playful,' while the minor mode is 'sad, dreamy and sentimental.'
(2) Firm rhythms are 'vigorous and dignified,' while flowing rhythms are 'happy, graceful, dreamy and tender.'
(3) Complex dissonant harmonies are 'exciting, agitated, vigorous, and inclined toward sadness,' while simple consonant harmonies are 'happy, graceful, serene, and lyrical.'
(4) Differences in expressiveness between rising and falling melodic contours are not clearly differentiated or consistent.

Subsequent studies have supported Hevner's finding of high level of agreement among listeners' characterizations of the general mood or emotion expressed by music (Gabrielsson & Lindstrom, 2001). For instance, Terwogt and van Grinsven (1991) studied children's and adults' ability to link musical excerpts to one of the following mood states: 'happiness,' 'sadness,' 'anger,' and 'fear.' In a procedure designed to accommodate children, participants viewed drawings of faces depicting the four emotions and pointed to the face that 'matched the music most closely' (p. 104). Consensus of reports was present even among 5-year-olds, but increased with age. Happiness and sadness were clearly identified, while fear and anger were not so well differentiated. The finding that some emotions may be more difficult to identify in music than others was also observed in unpublished research by W. G. Parrott and one of us (RH) on identifying the emotions expressed by various arias from Mozart's *Magic Flute*. Papageno's song in the first act was not only recognized as expressing happiness but participants also reported appropriate bodily feelings. On the other hand, the Queen of the Night's aria expressing jealous rage (according to the plot of the opera) was recognized as expressing some strong emotion but participants had no idea which strong emotion it was.

Listeners also seem to be able to detect some emotions in music from other cultures. For instance, Balkwill and Thompson (1999) played excerpts of Hindustani ragas to Western (Canadian) undergraduate students. Participants were asked to rate the ragas on the degree to which they expressed 'joy,' 'sadness,' 'anger,' and 'peace.' The Western students' ratings correlated strongly with the intended emotions of the ragas for three of the emotions: joy, sadness, and anger. In another study, Balkwill, Thompson, and Matsunaga (2004) also found that Japanese listeners recognized the emotions that Japanese, Western, and Hindustani music was intended to convey – namely, joy, anger, or happiness. In both studies, Laura-Lee Balkwill and colleagues observed that listeners' judgments of emotions in music corresponded to judgments of musical dimensions such as tempo, loudness, and complexity. For instance, music receiving high ratings for joy was usually judged as fast in tempo and low in melodic complexity, whereas music judged to convey anger was judged to be high in loudness and complexity.

Although there are not many cross-cultural studies of this sort yet, the findings thus far suggest that the ability to recognize emotions in music is not entirely dependent on the listener being familiar with the specific tonal system or particular conventions of a musical repertoire. For example, Dalla Bella and colleagues found that French-speaking children as young as 5 years old could distinguish the expression of happy versus sad in music, but only based on tempo (slow = sad, fast = happy), and not yet mode (major = happy, minor = sad) (Dalla Bella, Peretz, Rousseau, & Gosselin, 2001). Mode is a quality of music one learns through exposure to music, whereas tempo is linked to a quality that is general to all experience: velocity.

Hevner and other researchers have demonstrated that there are broad features of music (e.g., mode, harmony, tempo) that convey emotions with some degree of agreement among listeners; but can smaller-scale emotionally significant components of music be specified? The idea that there are elementary units of music that convey particular emotional meanings has been explored by musicologist Deryck Cooke (1959).

Specific tonal patterns

Cooke (1959) set out to show that the idea of 'music as a language capable of expressing certain very definite things' (p. xi) is not merely a fanciful notion. Proposing that the way emotion is expressed through music can be explained through the natural principles of physics (for instance, the degree of tension resulting from the relationship between the frequencies of two pitches), he charted the emotions conveyed by specific intervals of the diatonic scale. For instance, the idea that a major third may be heard as expressing joy, while a minor second conveys 'spiritless anguish' (p. 89) may be explained by the fact that the major third occurs early in the harmonic series while the minor second appears much later and has a more distant relationship to nature's harmony (see chapter 2).

This work culminated in Cooke's identification of some basic melodic patterns that are commonly found in Western tonal music (such as a stepwise movement from the tonic to the minor third and back again to the tonic, or C D E♭ D C if in C minor) and the corresponding emotion or emotions they convey (in this case, 'a trapped fear, or a sense of inescapable doom, especially when repeated over and over' [p. 140]). To illustrate and support his ideas, he provided many examples of vocal music in which the lyrics were shown to be consistent with the emotion he described. For instance, the minor third arch described here is repeated in the bass part of the 'Confutatis Maledictis' in Mozart's *Requiem Mass* and Violetta's repeated fervent plea 'you don't know what love I have for him' in Verdi's *La Traviata*, among a host of other examples provided by Cooke (pp. 140–143).

Cooke's conclusion takes for granted that everyone who can appreciate a certain system of sounds as Western tonal music must share to a greater or lesser degree in a common body of knowledge as to the emotional meaning of specific melodic contours as well. However, while the many illustrative examples of musical motifs that convey the same emotions are certainly intriguing, we do not yet have evidence that musical meaning is intrinsic. It is possible that music is inherently neutral and conveys no specific emotions in itself. Certain musical conventions for conveying emotion may have developed over time among composers, who continue these traditions. Listeners may learn these conventions through habitual music-lyric pairings in vocal and program music that eventually generalize to instrumental music. The question of whether music has intrinsic meaning, rooted in physical principles, remains an interesting proposition and has proved to be quite difficult to put to empirical test (cf. Gabriel, 1978).

How performers communicate emotion

The examples of how emotions are expressed thus far have focused solely on elements that can be seen in a notated score. However, an important characteristic of music performance is that the performer can deviate from the notated score in ways that communicate a certain expressed emotion, and this emotion may differ strikingly from the emotion that would seem to be conveyed by the notated score. In chapter 11 we discussed the ways in which performances can be expressive; here we focus on the link between performance expression and communicated emotion. Borrowing an earlier theory from vision, the Brunswickian 'lens' model, Patrik Juslin (1997, 2000) proposed that the expression of emotion from performer to perceiver involves the manipulation of a set of independent expressive 'cues' by the performer. Figure 14.2 shows a schematic of Juslin's 'lens' model of emotional

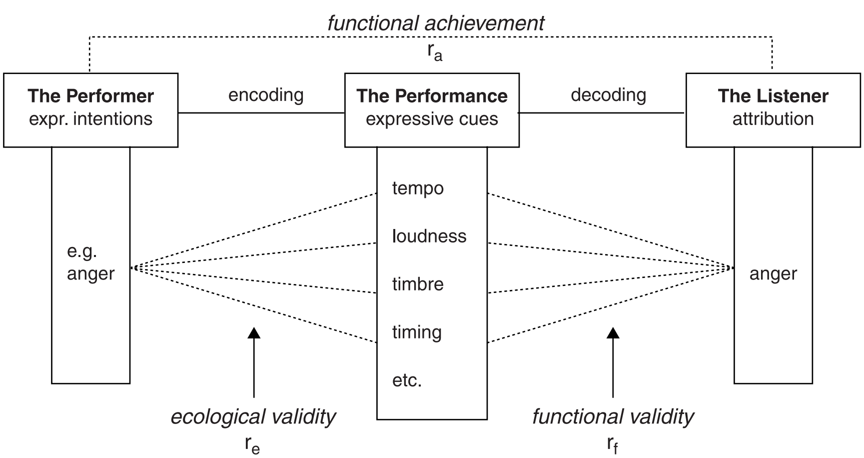


Figure 14.2 'Lens' model of emotional expression (Juslin, 1997, p. 394).

Source: Reprinted with permission from the University of California Press.

expression in music. The task of the performer is thus to take an intended emotion and to 'encode' the emotion using a set of appropriate values across all cues. Thus, a performer may choose to slow down or speed up tempo, may choose to make a note louder or softer, and so on. The listener, when presented with these cues, can recognize an emotion by 'decoding' the cues. Even when performers differ in how they interpret a particular piece, for instance when adding ornamentation (elaborations on the melody) in their own way, listeners are generally successful at recognizing intended emotions (Timmers & Ashley, 2007). Note that this model says nothing about whether an emotion is actually induced – again the focus here is on the expression of emotion.

So what cues do performers use? In summarizing the results of various studies, Juslin (2001) adopted the 'circumplex' model of emotion (Russell, 1980), which plots emotions across two dimensions – arousal (relating to the amount of energy in a musical passage) and valence (relating to expressing positivity versus negativity). A simplified version of his figure is shown in Figure 14.3. The upper right quadrant summarizes cues associated with high arousal and positive valence – i.e., 'joy.' The upper left quadrant is also positive in valence but peaceful – these are calm emotions. To the left of Figure 14.3 we see performance cues associated with negativity. In general the cues a performer uses reflects the kind of physical activity a person would

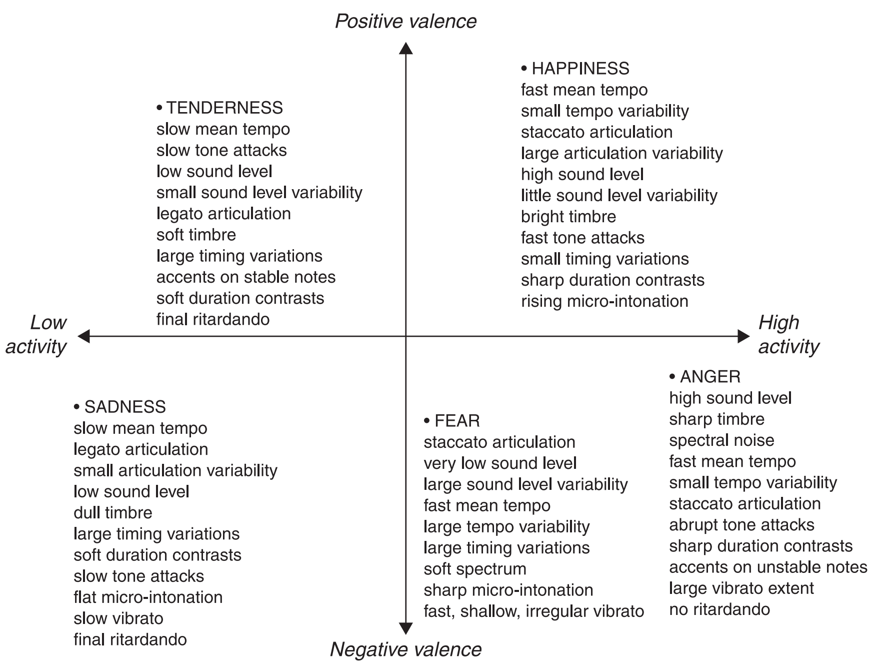


Figure 14.3 Examples of performance cues used to express five basic emotions (Juslin, 2001, p. 315).

Source: Reprinted with permission from the Oxford University Press.

engage in when feeling that emotion. Thus, happiness is associated with high energy but a quick (staccato) touch, sorrow (lower left) is associated with a kind of musical lassitude, as realized by a tendency toward slow and quiet performances. This similarity in the expression of emotion across musical and nonmusical contexts was further elaborated by Juslin and Laukka (2003) who suggest that there is a general connection between expression of emotion by the voice and expression of emotion in music. Or, in their words, 'music performers are able to communicate basic emotions to listeners by using a nonverbal code that derives from vocal expression of emotion' (p. 775). By this account, the expression of happiness in a performance is associated with greater energy than a sad performance because a person experiencing happiness tends to feel more energetic and would be likely to produce vocalizations that bear out this state.

Felt emotion in music

The research summarized above suggests that music may express or represent emotion in whole musical works as well as through specific musical features. But can music also *induce* moods or emotions in the listener? And if music is able to bring about emotions, do listeners experience the same moods or emotions – or do these 'felt emotions' vary widely between individuals? This fundamental question links with our discussion in our previous chapter on the emotivist-cognitivist problem. The idea that music expresses or represents emotions but does not elicit real emotions is consistent with the *cognitivist* position, while the view that music induces emotions in listeners has been characterized as the *emotivist* position (Kivy, 1990).

Self-reports of physiological responses

A study carried out some years ago by psychologist John Sloboda (1991) and reviewed in a more recent collection of papers (Sloboda, 2005) examined the relationship between structural and dynamic properties of music and reports of physical reactions germane to emotional experience. Participants were asked to recall pieces of music to which they remember having experienced a physical-emotional reaction (such as tears or shivering) upon listening to it, and to identify specific parts of the music that provoked these reactions. Such reactions are considered to be part of the 'physical expression' aspect of emotions, according to the standard psychological definition given earlier. Three groups of physical responses to music were identified as 'tears' (e.g., lump in the throat, crying), 'shivers' (e.g., chills down the spine, goose bumps) and 'heart' (e.g., a racing heart).

Sloboda then painstakingly collected musical transcriptions of the pieces of music the participants had spontaneously recalled in response to the task, identified the specific location and feature of the music that provoked the emotional reaction, and drew links between the specific features in the music and the recalled physical-emotional responses. The aims of the study were similar to Cooke's study discussed previously, but the focus was on emotions induced, as opposed to expressed, by the music – and the design of Sloboda's study was empirical.

'Tears' and 'shivers' responses were by far the most commonly reported physical manifestations, whereas 'heart' responses were more rarely reported. 'Tears' were most often provoked by melodic appoggiaturas and certain melodic or harmonic sequences; 'shivers' were mainly provoked by a new or unprepared harmony; and 'heart' responses by such musical features as a sudden dynamic or textural change or an event occurring earlier than anticipated. Sloboda has also amplified the scope of this study in a subsequent investigation of reports of bodily reactions to jazz and popular music (unpublished findings, reported in Sloboda, 2005, p. 210). Both self-report studies suggest that listeners do indeed have strong emotional and physical responses to music, and that the musical elements that seem to trigger these responses can be identified with some degree of consensus among listeners.

Sloboda delved into some detail in describing the structural and dynamic properties of the music that provoked physical reactions: 'What seems to characterize [the opening passage of Albinoni's *Adagio for Strings*] and other "tears"-provoking passages, are the successions of harmonic tensions or dissonances which are created and resolved within a structure which has a high degree of repetition and implication-realization' (pp. 210–211). As noted above, the structural dynamics of the passages that produce 'shivers' are quite subtle. Sloboda suggests that the key factor is that an event is expected at one level, while it is unexpected at another. Here Meyer's (1956) general thesis relating violation of expectations to emotional expressiveness (as discussed in the previous chapter), and the predictions of Narmour's Implication-Realization model (described in chapter 7) seem to fit the results.

The question of the generalizability of these findings across different styles of music has yet to be determined. In the unpublished study by Sloboda mentioned earlier, listeners tended to give combined physical reactions (e.g., reporting that they felt two or all three of the emotions simultaneously) to popular and jazz music. Perhaps more importantly, as far as we are aware, Sloboda's (1991) study has not yet been extended to other musical cultures. The question of whether these specific reactions are to be found in those who listen to the gamelan music of Java or the raga-based music of India is still to be investigated, so far as we know. While other studies (such as research by Balkwill and colleagues, discussed earlier in this chapter) have shown that listeners can identify emotions broadly in unfamiliar music, we do not know whether the links between particular musical features and specific emotional responses would be similar to Sloboda's findings. We would not necessarily expect the actual contours and other characteristics of emotion-inducing music to be the same in all musical cultures. In an informal classroom demonstration of Indian musical performances which were supposed to have emotional effects, one of the authors (RH) found that listeners moved by

Sloboda's examples had no coherent reactions to the playing of a raga by Vilayat Khan.

Peripheral physiological responses

Sloboda's study was based on self-reports of intense responses to music, but is there objective evidence that listeners actually *experience* these emotional responses? One phenomenon that has been shown in several studies to correspond with specific physiological changes is the 'chill' response which refers to 'a sudden, arousing reaction that is accompanied by goose bumps, shivers, or tingles in the spine' (Guhn, Hamm, & Zentner, 2007, p. 473), similar to Sloboda's 'shivers' cluster. For instance, in Guhn and colleagues' study, reported chills coincided with particular patterns of heart rate and increases of skin conductance. An analysis of the musical passages that tended to elicit chills showed that they shared some common features: Chills tended to occur during slow movements, when a solo instrument emerged or became distinct from the accompaniment, when there was an increase in loudness or a swell in the music, often also an expansion of pitch range, and when the harmonic progression was unusual or unexpected.

What about more specific felt emotions – do our bodies undergo physiological changes when listening, for instance, to music that induces feelings of happiness or sadness? To address this question, we turn to psychologist Carol Krumhansl's (1997) research, which compared self-reports of felt emotion with peripheral nervous system responses to music.

In Krumhansl's study, six musical selections were presented to participants: These comprised the three opening minutes of performances of compositions by Gustav Holst (*Mars* from *The Planets*), Antonio Vivaldi (*La Primavera* from *The Four Seasons*), Tomaso Albinoni (*Adagio in G Minor*), Modest Mussorgsky (*Night on Bald Mountain*), Samuel Barber (*Adagio for Strings, Opus 11*), and Hugo Alfvén (*Midsommarvarka*). One group of participants was asked to indicate the degree of sadness, fear, happiness, and tension they were experiencing as they listened to the music, by adjusting a slider on a computer. After listening to each excerpt, they were also asked to rate how they felt while listening to the music on a set of 13 scales representing emotions such as 'afraid,' 'angry,' 'anxious,' and 'surprised.' These 13 emotion dimensions were selected as they were relevant to the three basic emotions (sad, fearful, and happy) to which the study was directed.

The other group of students listened to the same six musical excerpts while their physiological responses were recorded by polygraph. Among the measures used were heart rate, blood flow, breathing, and temperature and dampness of skin; changes in these personal bodily attributes are the usual physiological responses to situations in which a person experiences one of the 'somatic' emotions. After each excerpt, they also completed the set of 13 rating scales.

The measures of felt emotions taken during listening revealed that the six musical selections induced three distinct emotions: The Albinoni and Barber excerpts yielded high ratings for sadness, the Holst and Mussorgsky for fear, and the Vivaldi and Alfvén for happiness. (The ratings for tension correlated with all three emotions, which suggests that tension may be influenced by each of these emotions.) In general, the physiological readings also differed systematically for these pairs of excerpts, indicating emotion-specific physiological responses corresponding to the emotion of the music. The two measures that were taken *while* listening to the music (emotion judgments indicated by the slider and the physiological measures) were also correlated with each other, although the magnitude of the correlations was small. As both of these measures were taken on a second-by-second basis for the duration of each excerpt, the correlations reflect the correspondence between movements of the slider and changes in the physiological measures. In general, slider movements along the 'sadness' dimension corresponded with changes in cardiac and electrodermal systems, 'fear' dimension with cardiovascular changes, and 'happiness' dimension with changes in respiration measures (p. 349).

Krumhansl's (1997) findings show that 'not only do listeners verbally report emotional responses to music with considerable consistency, music also produces physiological changes that correspond with the type of musical emotion' (pp. 350–351). She concluded that 'the physiological effects of music observed [in this study] generally support the *emotivist* view of musical emotions' (p. 336). The findings also point to the dynamic nature of musical emotions. Few correlations were found between averaged physiological measures and the ratings on the 13 self-report scales taken *after* each excerpt was played, suggesting that musical emotion may not be well captured in a single 'snapshot' reading after the fact. Its dynamic and temporal character may be more faithfully represented in measures taken across time as the music unfolds.

Neural bases of emotional responses to music

Krumhansl's study showed that listening to music can induce emotions which correspond with specific physiological changes in the body. What, if anything, happens in the brain as a mood or emotion state is elicited by music?

There is as yet not much neuroscientific research on emotions induced by music. One line of research pursues one of the reasons suggested for *why* we listen to music – namely, the idea that music offers some kind of intrinsic reward for our nervous system, like a piece of cheesecake or a roller-coaster ride (cf. Huron, 2003; Pinker, 1997). These statements might seem to trivialize the meaningful experience many of us have of music, but nevertheless the 'appetite' of our current population for music suggests that music listening has an almost addictive quality (Huron, 2003; Levitin, 2006). People often report that listening to music makes them feel good, lifts their mood, and can trigger emotional reactions of unusual intensity and even transcendent experiences (see Gabrielsson, 2001, on 'strong experiences' with music). Not surprisingly, music appears to have an impact on the brain's reward centers. For instance, evidence from an fMRI study suggests that *consonant* musical intervals (which are pleasant and harmonious to the ear) stimulate a region of the orbitofrontal cortex (interior relative to the region shown in Figure 4.2), which is associated with reward and reinforcement (Blood, Zatorre, Bermudez, & Evans, 1999; see also Mitterschiffthaler, Fu, Dalton, Andrew, and Williams, 2007 for a similar result with longer musical excerpts). By contrast, *dissonant* intervals increase activity in the parahippocampal gyrus, a region that has intricate connections to the amygdala, the brain's 'warning center.' Further evidence of the role of the amygdala as a warning center extending to music comes from a study showing that removal of the amygdala (to prevent seizure) leads to a reduced ability to recognize 'scary' music (Gosselin et al., 2005).

Interestingly, the data yielded by the study by Blood and colleagues (1999) did not show any difference in the responsiveness of the auditory cortex to consonant or dissonant music. Could it be that differentiating musical emotions does not involve the use of our primary auditory system? If the auditory system does not respond to musical differences that influence emotional response, then perhaps there is a dissociation between our ability to experience emotions and our ability to perceive music. Although cognition and emotion have traditionally been separated in psychology, more recent thinking tends to consider these functions as interdependent (e.g., Damasio, 1994). One might even consider the ability to perceive music to be a prerequisite for emotion.

Contrary to this intuition, however, Isabelle Peretz and colleagues (Peretz & Gagnon, 1999; Peretz, Gagnon, & Bouchard, 1998) have identified cases in which brain damage disrupts the ability to perceive and remember music, but does not disrupt the ability to identify the emotion that music is intended to communicate. For example, upon hearing the melody 'Happy Birthday' without lyrics, patient IR said 'I don't know that tune but it sounds happy' (quoted in Peretz, 2001, p. 116). The two individuals in the cases reported by Peretz and colleagues (Peretz & Gagnon, 1999; Peretz et al., 1998) had suffered severe damage to the auditory cortex, but not to orbitofrontal regions thought to mediate emotional responses (as in Blood et al., 1999). In each case, the individual suffered severe deficits in the recognition of familiar melodies and in detecting changes to melodies. However, both persons correctly labeled the emotion being communicated by music, and appeared to experience that emotion themselves.

Application: The emotional power of music in film

We conclude this chapter by applying the topic of emotion in music to a real-world context that has general relevance in our lives: the role of music in film. In the context of film music, not only is an external idea supplied as in the imaginative title or synopsis in program music, but concrete visual images accompany the music within the context of a story line and fictional characters. Under these conditions, does music still carry emotional weight and meaning of its own, or is the emotional meaning driven primarily by the powerful moving image?

Music has always played an important role in motion pictures. Even 'silent' films were not silent; before the advent of sound films, music supplied by live musicians masked the sound of noisy projectors and restless audiences, and broadly underscored the general mood and actions shown on screen. Since the introduction of sound films in the 1920s, the soundtrack has played an increasingly important role in contributing to the immersive quality of film. Film sound is not contained in a flat, two-dimensional screen of a prescribed size, as are film images (Chion, 1994). Thus, sound adds 'a third dimension' to the film experience as the audience is immersed in an 'envelope of sonority.' The viewers gaze at a rectangular screen to which most conscious attention is directed, but are surrounded on all sides by sound. While watching a character fleeing from imminent threat during a heavy downpour, as viewers we hear the sinister heavy breathing and accelerating footsteps following right behind us, while the roar of rain showers surround us on all sides. Sound thus places the viewer *in the center* of the experience, rather than as a spectator (Doane, 1980).

Music has the power to further immerse the viewer by heightening emotions. It is no surprise, therefore, that the research on film music has largely focused on its emotional impact (Cohen, 2001). While the 'cognitivist-emotivist' problem has also emerged as a topic of debate in the discourse on film music (see Smith, 1999), it is generally assumed that film music can both express and induce emotion. Both effects may heighten the audience's attention to onscreen events. In a dramatic demonstration of emotion induction, for instance, Thayer and Levenson (1983) found that physiological responses to a graphic film about industrial accidents were significantly stronger when the film was paired with a 'horror' soundtrack than with more neutral 'documentary' music.

Music that expresses emotion has also been shown to intensify the perception of the emotional content of images. For example, Bolivar, Cohen, and Fentress (1994) paired 'aggressive' or 'friendly' music with filmed interactions of (real) wolves that were either friendly or aggressive. Overall, the aggressive interactions were rated by viewers as significantly more aggressive than the friendly interactions. However, within the aggressive and friendly interactions, the music either increased or reduced the perceived degree of aggressiveness or friendliness in ways that were congruent with the emotion expressed by the music.

The soundtrack and images do not always mirror each other with respect to emotion or mood. Often, the relationship between sound and images in film is 'empathetic' (congruent with the emotions portrayed on screen); a hot pursuit is usually accompanied by the 'chase music' that the audience has come to expect. In other cases, the soundtrack may be 'anempathetic' – or seemingly indifferent – to the emotion or mood of the scene. After the horrific murder takes place in the classic shower scene in Alfred Hitchcock's *Psycho* (1960), the sound of the running shower continues 'as if nothing had happened' (Chion, 1994, p. 9). As Chion has observed, anempathetic music can create the effect 'not of freezing emotion but rather intensifying it' (p. 8). In some cases, the music and scene may be sharply juxtaposed. In Steven Spielberg's *Minority Report* (2002), two characters make their way through a shopping mall in a rather tense chase scene, accompanied by the gentle song 'Moon River' playing over the speaker system inside the mall. (We shall return to this film again later in this discussion.)

Boltz (2004) has proposed that information from images and music that are juxtaposed in mood may be encoded separately, as opposed to being processed in an integrated fashion. She found that memory for the details about a film scene and the accompanying music were significantly poorer when the images and music were mood-incongruent than when they were matched for mood. Further, when viewers were instructed to focus only on the images or the music, incidental learning occurred in the other mode but only for mood-congruent films. It is possible that mood-congruent information provided in more than one sensory modality is more likely to be integrated into memory as a unified whole than cross-modal information that is mood-incongruent.

Film music can also wind up tension and heighten emotions by subtly influencing the audience's interpretation of the unfolding storyline. Vitouch (2001) showed a fairly neutral and open-ended scene from Billy Wilder's film *The Lost Weekend* (1945) with one of two soundtracks, and asked participants to describe what they thought would happen to the characters after the excerpt ended. Viewers were more likely to give positive or ambivalent story continuations if the scene was accompanied by the (original) Rózsa soundtrack, and more negative story continuations if shown with Barber's *Adagio for Strings*, a piece found in previous research (including Krumhansl's 1997 study described earlier) to induce 'sad' or 'melancholic' emotions. As Cohen (2008) has proposed, 'the audience is constantly synthesizing a story, a *working narrative*, derived jointly from sensory information and hypotheses or expectations based on long-term memory' (p. 443), and 'cues from music as well as from the other sources of information contribute to the working narrative' (p. 444). A full explanation of other sources leading to the working narrative can be found in Cohen's discussion of her congruence-associationist model (Cohen, 2008; see also Marshall & Cohen, 1988, for the original formulation of the model).

Different soundtracks may also lead to different assumptions about more subtle details within this working narrative, such as the intentions of a character or the nature of the relationship between characters (e.g., Boltz, 2001; Bullerjahn and Güldenring, 1994). In Boltz's (2001) study, for instance, participants who saw a scene accompanied by music conveying a 'positive' mood were most likely to interpret the relationship between two characters in a scene as harmonious or romantic. Participants viewing the same scene with music conveying a 'negative' mood, on the other hand, were more likely to infer that one character would harm the other. Positive music also led to more positive descriptions of the male character's traits (e.g., kind, loving, protective), while more negative personality descriptions were ascribed to the male character (e.g., deranged, evil, manipulative) when the same scene was accompanied by negative music. Similar results have been yielded in studies investigating the effect of different musical soundtracks on the interpretation of the behavior of animals (Bolivar et al., 1994, discussed earlier) and even the attributes of geometric shapes (Marshall & Cohen, 1998) in short films.

While many studies have shown that changing the soundtrack can affect perceptions of onscreen images, more subtle manipulations of the soundtrack can also have surprising effects. For instance, Tan, Spackman, and Wakefield (2008) examined the effects of presenting the *same* piece of music diegetically and nondiegetically. The *diegesis* refers to 'all that belongs . . . to the world supposed or proposed by the film's fiction' (Souriau, quoted in Gorbman, 1987). Thus a fast music soundtrack accompanying a car chase would be *nondiegetic*, while a juke-box playing in the background during a bar-room brawl would be *diegetic* as it is supposed to exist within the fictional world inhabited by the characters.

Siu-Lan Tan and colleagues (2008) selected the scene described earlier from Spielberg's *Minority Report* (2002), showing a man and woman moving hurriedly through a shopping mall, accompanied by an orchestral rendition of Henry Mancini's 'Moon River.' As shown in Figure 14.4, the relationship between the two characters and their situation is rather open to interpretation for those who have not seen the film. Indeed, the findings showed that participants interpreted the scene differently – depending on whether the music was presented diegetically or nondiegetically. Those who watched the *diegetic* music version (music coming from inside the mall) judged the scene to be more tense and suspenseful, and perceived the relationship between the two characters to be more antagonistic, hostile, and ill-intentioned than those who watched the nondiegetic music version. On the other hand, participants who saw the *nondiegetic* version (consisting of the *same music* but presented as a dramatic score, external to the world of the characters) perceived the male as experiencing less fear, less excitement, and less romantic interest in the other character. It is possible that Mancini's gentle ballad 'Moon River' was perceived as incidental to the scene when it sounded like mall music playing within the environment of the characters, but as a commentary on the situation unfolding on screen when presented as the dramatic score. The participants had been told that the aim of the study was to examine how viewers interpret film characters' emotions, so the effect occurred even when they were mainly attending to the screen images as they would in a theatre.

In the Tan et al. (2008) *Minority Report* study, a third version of the same

Figure 14.4 Six images from *Minority Report* (DreamWorks and Twentieth Century Fox, 2002), directed by Steven Spielberg and produced by Gary Goldman and Ronald Shusett, actors are Tom Cruise and Samantha Morton. This excerpt was used in Tan, Spackman, and Wakefield's (2008) study.

Source: Please see reference list under director's name for information.

film sequence was also created, using 'chase music' from a John Williams film score accompanying a tense chase scene. At the end of the study (when it was finally revealed that the soundtracks had been altered in some conditions), participants were shown all three versions. Most of the 245 participants thought the 'chase music' was the original music that accompanied the scene – most likely as it was congruent with the apparent mood of the screen images. Interestingly, only if Spielberg's original (diegetic 'Moon River') version was shown first were participants more likely to correctly identify it as the one used in the actual film. It is notable that although most participants thought the 'chase music' originally accompanied the scene, Spielberg's choice of music (the diegetic 'Moon River' soundtrack) yielded the highest tension ratings for both mood of the scene and relationship between the characters – suggesting a masterful director's intuition for music that might not be the expected choice but would most intensify the scene!

Music may not even need to be presented simultaneously with a scene in order to influence viewers' interpretations. In another study, Tan, Spackman, and Bezdek (2007) paired film excerpts that featured a solitary character (showing no strong emotion) with nondiegetic music that had been previously shown to convey 'happiness,' 'sadness,' 'fear,' or 'anger.' (Still frames of a few of the film excerpts are shown in Figure 14.5.) The music was played *before* or *after* a solitary character entered and exited the scene, while lingering shots of the character's face and full body were shown without music. When asked to choose their own words to describe the character's emotion, participants tended to provide emotion words that corresponded with the emotion of the music – even though the music was not played simultaneously with the character. When asked to complete rating scales to convey the character's emotions, participants' ratings reflected a somewhat more complex mixture of emotions, but again the highest ratings for the 'sadness,' 'fear,' and 'happiness' scales were yielded for the films with the corresponding music-emotion soundtrack. The Tan et al. study showed that music not only influences interpretation of simultaneously shown images, but may even influence interpretations of previous and forthcoming scenes (see also Boltz, Schulkind, & Kantra, 1991).

In some cases, music may convey something of emotional import that is not shown on the screen. For instance, past and future events not present in the accompanying scene can be represented in film through the use of *leitmotifs*, a term borrowed from Wagner's operas, referring to a recurring theme that through association comes to represent a character, idea, or event throughout a work. The skillful use of leitmotifs is well illustrated in John Williams' original score for *Star Wars* (Lucas, 1977), a film that has aptly been referred to as an epic 'space opera.' Leitmotifs for Darth Vader, Luke, Leia, and Obi-Wan Kenobi, among others, are introduced and deftly interwoven with each other into the rich orchestral score, often at emotional high points. In one scene, for instance, young Luke watches the double sunset at Tatooine. In composer John Williams' words:

Figure 14.5 Images from three films used in Tan, Spackman, Bezdek's (2007) study: *Swimming Pool* (Fidélité, 2003), directed by François Ozon, produced by Olivier Delbosc and Marc Missonier, actor is Charlotte Rampling; *Three Colors: Blue* (CAB Productions, 1993), directed by Krzysztof Kieslowski, produced by Marin Karmitz, actor is Juliette Binoche; and *Diva* (Les Films Galaxie, 1981), directed by Jean-Jacques Beineix, produced by Claudie Ossard, actor is Wilhelmenia Wiggins Fernandez.

Sources: Please see reference list under director's names for information. This figure appeared in *Music Perception*, *25*, p. 140, reprinted with permission of University of California Press.

> Originally, I scored the scene with Luke's theme. When George Lucas heard it, he asked if I could replace it with Ben's theme [Obi-Wan Kenobi]. George's feeling was that since Luke dreamed of leaving Tatooine and becoming an adventurous spacepilot [like Ben], Ben's theme is better in that context.
>
> (Quoted in Kendall, 1993)

Luke's thoughts and emotions are subtly suggested through the music, transforming a quiet, reflective scene into a more emotionally powerful one. While leitmotifs are usually woven into the dramatic (nondiegetic) score, a notable use of a diegetic leitmotif occurs in *Pirates of the Caribbean: Dead Man's Chest* (Verbinski, 2006), in which Davy Jones plays a pipe organ in a climactic destruction scene. The emotional resonance of this scene is intensified as the theme he performs in this grand gesture is his own leitmotif, which was played throughout the film on a musical locket. Lexmann (2006, p. 112) points to an interesting example of the use of a nondiegetic motif. Although nondiegetic sounds are not supposed to be heard or sensed by the characters as they do not exist in their fictional world, the main character in Robbe-Grillet's 1968 film *The Man Who Lies* is apparently frightened of the recurring sound of a small drum in the dramatic score, responding as if it conjures up the image of a machine gun. Here, the interplay between diegetic and nondiegetic is surprising, and suggests a character who may be alternating between the real and the imagined.

Musical tension-resolution in the musical soundtrack, as discussed in the theories of musical structure in chapter 7 and Meyer's theory in chapter 13, may also have an impact on the perception of the structure of the film. For example, a study by William Forde Thompson and colleagues showed that film excerpts are judged to have ended with greater closure if accompanied by a musical soundtrack that resolves to the tonic (e.g., ending on a C chord if in the key of C) than if the music remains unresolved (Thompson, Russo, & Sinclair, 1994). Further, the effects are stronger for tonally closed music with a clear metrical structure and ending on a strong beat. A striking example can be found in the second and third films in the Wachowski brothers' *The Matrix* trilogy, scored by Don Davis. The final scene of *The Matrix Reloaded* (2003) ends with an unresolved progression of dramatic chords that leaves the audience hanging, whereas its sequel *The Matrix Revolutions* (2003) ends with a dominant-to-tonic cadence that gives the final scene a strong sense of closure. There are also films in which a new piece of music begins during the last moments of the film and proceeds seamlessly through the end credits (all the films in *The Bourne* trilogy: Liman, 2002; Greengrass, 2004, 2007), and those in which the music and closing scene end or fade together (the American flag in *Saving Private Ryan*; Spielberg, 1998) giving a stronger impression of finality. Thus the role of music on the emotional and dramatic impact of a film is evident up to its final closing scenes.

As we have seen, music provides a rich conduit of emotional communication to the film audience, apart from the information contained in the images, dialogue, and other sounds. Film music is not relegated to the role of simply mirroring or intensifying the effects of the screen images, but can contribute meaning of its own that advances the narrative, juxtaposes with the scene, reframes our perceptions of preceding scenes and creates expectations about upcoming scenes, conjures up ideas and images that may not be present on the screen, and signals whether the story is resolved or remains open-ended. Although viewers often do not consciously attend to the 'unheard melodies' of cinema (Gorbman, 1987), the emotional force of the film score – in part through its power to convey and induce emotion – influences the interpretation of images and unfolding narrative in subtle but powerful ways.

Coda

What musical features of Simon and Garfunkel's 'Bridge Over Troubled Water' are the grounds of a judgment on the emotional content of the piece? What does it mean to be emotionally 'moved' by this song – does it entail an intellectual recognition of the emotions represented in the music? Does it also involve a bodily emotional response? What emotional impact would this song have on someone who does not understand the words? This chapter examined the emotional significance of music. From a convergence of empirical studies employing many methods including self-report, physiological, and neuroscientific methods, we have observed the profound impact of music on emotion. Even when attention is shared among several sensory modalities as in the viewing of an opera or film, the emotional power of music in its ability to represent, express, and induce emotion is clearly manifested. Still, questions remain concerning both the way in which music induces emotion, and the degree to which music's effect is emotional in the classic psychological sense. For instance, the mediating effects of things we associate with music from our past experience have not fully been distinguished from the effects of music itself (cf. Konečni, 2008).

One issue that emerges from these questions is whether the emotional significance of music for one cultural group persists when that music is presented to a different group. More generally, one might question whether the structural conventions that have emerged in Western music are shared among other cultures. To this point we have summarized a few studies in the present chapter that address musical experiences across cultures. In the next chapter we address the issue of culture more directly, by surveying the case for and against 'musical universals,' and by reviewing some of the research that has aimed to understand human musicality on a global scale.

15 Culture and music

We conclude our book with a topic that is often of immediate interest to students of music psychology but has been somewhat slower in taking root in formal research: To what degree does *everything we have discussed so far* depend on one's cultural context? Cognitive psychology, which provides part of the foundation for the research summarized here, tends to assume universalism. The components of thought are often referred to as 'mechanisms,' suggesting that the process of thought in all cultures is put into motion by a common set of bolts and gears (speaking metaphorically of course). A tacit assumption of universalism is typically adopted. With respect to musical experiences, universalism involves the assumption that – to some degree – all people from all cultures experience music in the same way. Thus, although the music of different cultures may sound strikingly different to various listeners, similar brain processes are used in their perception and thus we may expect that a common set of core principles may allow us to understand how people experience these differing sound patterns. At the same time, most cognitive psychologists would admit that culture plays some significant role in the actions, thoughts, and feelings that are involved in musical experience. Musicologists, as whole, typically place greater emphasis on cultural differences than do cognitive psychologists (Carterette & Kendall, 1999). This makes sense, given that the point of musicology is to better understand a product of human culture (music) rather than the generic mechanisms of behavior (psychology).

It is important to note that one rarely, if ever, would find a thinker so extreme as to entirely dismiss the contribution of culture. However, it is similarly shortsighted to assume that *every* aspect of musical experience is cultural, as some are clearly based on fairly universal properties of physiology (e.g., the recording of a dog whistle on 'Sergeant Pepper's Lonely Hearts Club Band' was recorded by Paul McCartney so that his dog, but not human listeners, would be entertained). It is also an error, as we noted in chapter 4, to assume that anything physiologically based is therefore universal, as almost every aspect of human physiology is subject to environmental influences.

The challenges of cross-cultural research

People are often surprised by the dearth of scientific cross-cultural research on music, which was quite sparse until the late 1990s. Why would a field lack research in an area of such self-evident importance? One reason may stem from researchers' awareness of their own limitations. Most music psychologists are Westerners, steeped in the European tradition of (primarily classical) music. When exploring a new topic, it makes sense to begin with a context that complements one's expertise.

A second reason may stem from the practical and methodological complexities of conducting cross-cultural research. Quite simply, it is difficult to find research participants whose musical experience solely reflects a non-Western culture. One can identify individuals from one's own surroundings who were raised in a different culture. For instance, in an aforementioned study by Hannon and Trehub (2005a, see chapter 6) the authors recruited participants from the Toronto area who had grown up in the Balkans. Such individuals are in short supply and also are effectively bicultural, having acquired musical instincts from two cultures. Another option would be to recruit in the country and region of the target population, as did Krumhansl and colleagues (2000) in their study of Sami Yoiks (see chapter 7). Even doing this, however, does not ensure isolation from Western music. For better or worse, musical styles are vanishing, probably more rapidly than are languages, as our world becomes a more homogeneous place. Accordingly, the scope of study in ethnomusicology has changed. Whereas before the 1970s, ethnomusicologists often neglected the study of music combining the styles of two or more cultural sources, there is now more fieldwork on music that reflects a cultural mix of musical styles (Nettl, 2001).

There is also the issue of cultural sensitivity. It is rare to encounter research on music cognition that adopts an anthropological perspective (though the field of 'cognitive ethonomusicology' is on the rise). One reason for this omission may arise from the respect researchers have for the inherent value of each culture's music. To many, 'universalist' perspectives are ultimately 'Westernized' perspectives, and theories that couch Western notions as 'universal' implicitly position a certain cultural context (that of Western Europe) at the pinnacle. As ethomusicologist Bruno Nettl (2001) has expressed it:

> We should . . . eliminate the notion that all musics pass through a set of stages, and that we can explain the variety of world musics by suggesting that we are observing each of them at a different stage of the same development . . . It is not inevitable (or even likely) that a non-Western music would gradually change to become like Western music. (p. 13)

This concern is not mere fantasy, of course, given that explicit claims to this effect have been made periodically through history.

With this understanding in mind, we proceed with all due modesty to discuss some of the existing cross-cultural research on the psychology of music. Some of this research has been discussed in earlier chapters. Our goal here is to bring this research into the broader context of cultural 'universalism.'

Music in nonhuman animals

It may be surprising to see a chapter on culture beginning with a discussion of research that does not focus on humans. However, one can argue that some of the strongest evidence for or against universalism should come from comparative research. Based on the logic of evolutionary biology, one might consider music-like traits that we share with closely related nonhuman animals to constitute core 'universal' musical characteristics that have developed over time as a result of evolutionary adaptations.

There is some evidence for shared musical faculties across species, though there are also differences in both results and interpretation. It is important to note at the outset that the prevalence of 'song' in birds may not be the best place to look. Although it has been suggested that vocal learning of songs in birds may be related to vocal learning (including music) in humans (Jarvis, 2004), birds are rather distantly related to humans. And, as we will discuss later, it is unclear whether the 'song' of birds is truly like songs that humans sing.

So we turn to more closely related ancestors: monkeys. (Great apes do not appear to have been studied yet with respect to music perception or cognition, possibly because their formidable strength, size, and aggressiveness make them more difficult to use as subjects.) Although the vocal physiology of nonhuman primates precludes their ability to sing (e.g., Mithen, 2006), their auditory capabilities may nonetheless show musical capabilities similar to ours. Impressive efforts have attempted to discern whether monkeys process basic musical features as do humans. One question that emerges, of course, is what exactly constitutes a 'basic musical feature'? Two possibilities have been addressed.

One is octave equivalence. As mentioned in chapter 5, *octave equivalence* is one way in which our perception of music differs from the basic psychoacoustics of sounds, in that tones separated by an octave may be treated as more similar (identical in chroma) than tones with frequencies that are more proximal in pitch height. Interestingly, many studies have not found evidence for octave generalization in monkeys (see McDermott & Hauser, 2005 for a review). One exception is a study by Wright and colleagues (2000), who conducted a series of experiments in which rhesus monkeys learned through reinforcement to distinguish various auditory stimuli, some of which were melodies. When presented with tonal melodies separated by an octave, monkeys behaved as if no change had occurred – thereby showing octave generalization. Importantly, this effect only worked for tonal melodies and only when constituent tones were complex tones rather than sine tones. This finding is impressive taken on its own, but does not sit well with many other findings according to research reviewed by McDermott and Hauser (2005), which generally suggest that nonhuman animals (including primates) experience music as a set of fixed pitch classes (i.e., in an 'absolute pitch' sense) rather than as pitch relationships ('relative pitch'). Thus one must say there is 'mixed' support for the idea that we inherited octave generalization from our ancestors.

Other studies have examined whether nonhuman primates are sensitive to another fundamental quality of music: *consonance* or *dissonance* of musical chords. McDermott and Hauser (2004) placed cotton-topped tamarins in a cage shaped like a 'V' as shown in Figure 15.1, configured so that a consonant chord sequence would sound if the monkeys traveled down one leg and a dissonant sequence would sound if they passed to the other leg. The tamarins acted as though it made no difference whether consonant or dissonant chord sequences were playing. They spent similar amounts of time in both sections, even though they displayed a marked preference for harsh unpleasant sounds versus silence. Thus these close ancestors may not be able to make sensory distinctions that are fundamental to musical categories in tonal systems. In fact, a similar subsequent study suggests that tamarins prefer complete silence to any kind of music (McDermott & Hauser, 2007). By contrast, humans placed in a similar context (e.g., a room configured so that consonant chords are played when walking to one side but dissonant chords are played

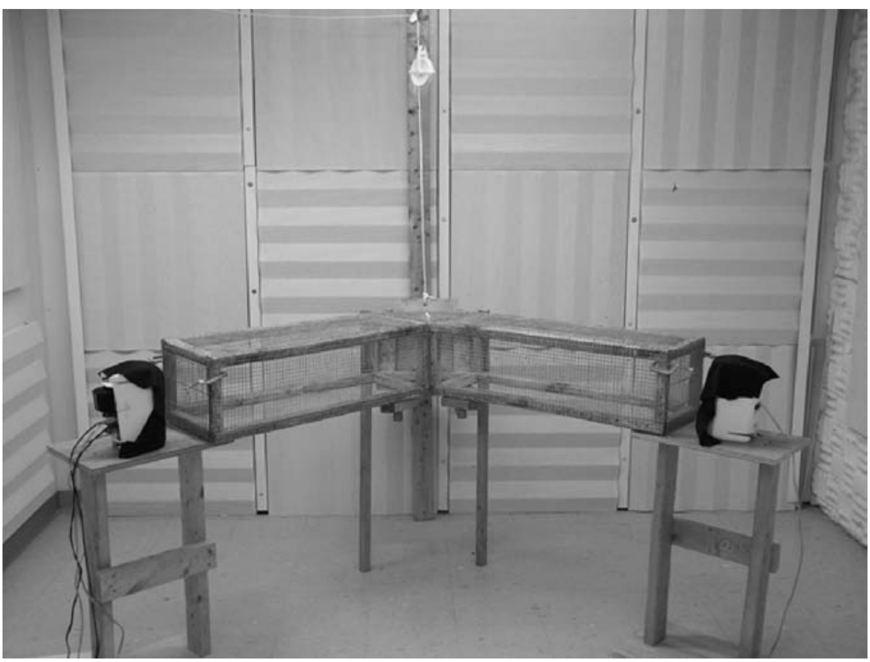


Figure 15.1 Apparatus used in McDermott and Hauser (2004, figure 1, p. B13).

Source: Reprinted with permission from Elsevier.

when walking to the other) show the kind of specific preferences one would expect, preferring consonance to dissonance and music to silence (McDermott & Hauser, 2004, 2007).

Thus it is not clear that our close ancestors possess comparable musical faculties to us, even for very 'basic' musical 'features.' It is, however, a logical leap to conclude from these data that music is uniquely human. First, it is not yet determined whether animals like chimpanzees (more closely related to humans than tamarins) differ from humans in the way they process musical features. Also, there is the presence of music-like behaviors in nonprimates, such as birdsong. Perhaps birds are a better prospect for seeking comparative universals than are primates. In particular, one important shared trait between humans and many (though certainly not all) birds is vocal learning. The capacity to acquire a broad repertoire of auditory sequences for production and perception, and the capacity to create novel combinations of these sequences, suggests an interesting commonality between humans and birds. Evolutionarily, such shared traits are thought to be *analogies*, similar traits that emerge in unrelated species due to environmental demands, rather than *homologies*, traits that are inherited from an immediate ancestor. Interestingly it has recently been demonstrated (as seen in examples on 'YouTube' and elsewhere) that birds who engage in vocal learning can also dance to music. For instance, Patel and colleagues have shown that the sulphur-crested cockatoo can synchronize to the beat of music, even adjusting movements (such as bobbing its head and kicking its feet) to match music played at different *tempi* or speeds (Patel, Iverson, Bregman, & Schultz, 2009).

Having said this, it is not clear whether birdsong is truly 'song' in the sense that we think of the word. McDermott and Hauser (2005) argue that animal 'songs' are far more fixed, both with respect to structure and function, than human musical song. The structure of birdsong, for instance, is far less varied than human music (at least with respect to our present abilities to define and analyze musical variety). Functionally, at least as far as we are aware, songs among animals have very specific roles and are typically used either to attract mates or to protect one's own domain. Certainly human music can take these roles, but we often create music or listen to music simply because it is pleasing. The musical babbling and spontaneous singing of infants, discussed in chapter 9 for instance, often has no immediate pragmatic function. In addition, it appears that the perception of pitch among birds is more constrained by absolute pitch than it is for humans; that is, birds are more likely to be 'fooled' by transpositions of melodies (Hulse, Cynx & Humpal, 1984).

Whether or not human and bird song are comparable is ultimately a complicated issue that we cannot fully consider here. An excellent selection of papers on this complex topic is offered by Wallin, Merker, and Brown (2000). For now, we simply point out that such apparently simple parallels are not so clear cut when one delves more deeply into the issue.

Music-making across cultures

We now address whether cultural universalism can be said to exist among humans. This issue can be addressed in two different ways: by examining the kinds of musical structures present in the musical traditions of different cultures, and the sensitivity of persons in different cultures to musical features present in each other's music. Both speak to the problem of universality in different ways. The kinds of structures that emerge in the culture speak to the ways in which people develop certain core patterns that constitute a kind of 'syntax' for that culture's musical 'language.' To the extent that these different structural traditions share common traits, we may consider the commonalities as part of a universal structure, potentially related to the auditory and motor systems. The second issue addresses a related point: Even when cultural traditions diverge, people may be able to comprehend the traditions of other cultures to the extent that a common musical faculty guides the processing of music across all cultures. In the discussion that follows, we will explore this problem by considering research on cultural differences in various facets of music perception and production.

Pitch and tonal structures

One of the notable differences that can be observed across musical traditions is in the expression of pitch. For instance, while Western orchestras take great pains to tune their instruments as closely as possible to a standard pitch, pitched instruments of the Indonesian *gamelan* orchestra are tuned about 7 Hz apart (*penyorog* tuning) to each other, creating a pulsing, shimmering quality (Hood, 1966). In much of Western music, each pitch is typically produced as a discrete unit, with abrupt changes between pitches. By contrast, pitches in Indian music are often allowed to waver, and a performer will often glide gradually from one pitch to the next. Western listeners can often find such wavering in pitch to be disconcerting at first. However, it is important to recognize that popular music in the West often features deviations in pitch from established 'standards,' as in the glides that are representative of blues or the use of melisma that is common among singers like Mariah Carey. Clearly, there are rich variations in the use of pitch both in formal musical systems and in expressive performances. Moreover, many cultures do not use the same set of pitches or scale structure to create melodies.

Despite these differences in the realization of musical pitch, Justus and Hutsler (2005) describe some core pitch features that may be held in common across cultures. Figure 15.2, taken from their paper, shows canonical scale structures used in the music of many cultures. The main commonality is that all cultures appear to incorporate octave equivalence in their tonal systems (though see also Tenzer, 2000, discussed in the next section). Also, many cultures adopt a core set of pitch classes (referred to as 'tonal material' in Figure 15.2), on which larger scale structures are built. These pitch classes

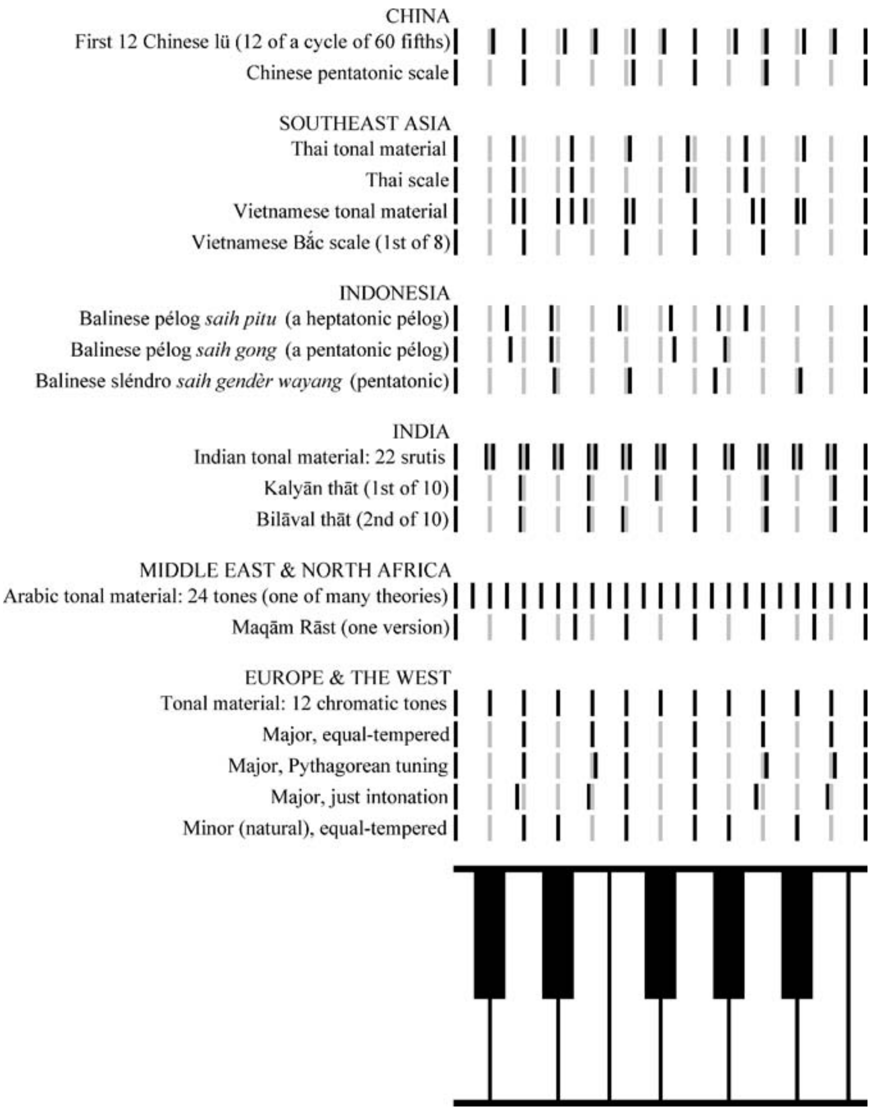


Figure 15.2 Justus and Hustler's (2005, figure 4, p. 13) representation of pitch in the tonal systems of various cultures (dark bars) as compared with the 12 chromatic pitch categories of the Western equal tempered scale (gray bars and piano keyboard).

Source: Reprinted with permission from the University of California Press.

appear to be equally spaced in most cultures but do not always adhere to our semitone scale (e.g., Thai tonal material).

Cultures also differ with respect to their use of so-called 'tonal material.' Whereas Western music largely relies on the diatonic scale, other cultures have adopted different conventions. However, in one other important respect, one can refer to a kind of universality in scale structure. Namely, every culture shown in Figure 15.2 adopts scales (from tonal material) for which intervals are spaced *asymmetrically*. That is, scales created from these materials tend to involve intervals with different spacing (e.g., the Western diatonic scale is constructed out of 'whole steps' (2 semitones) and 'half steps' (1 semitone). By contrast, tonal material (e.g., semitones in Western music) is often equally spaced. As such, Justus and Hutsler (2005) propose that asymmetry, though probably not diatonicity, may be culturally universal. There are also some commonalities of intervals present in tonal systems. Nettl (2001) has proposed that 'the principal melodic interval in most of the world's musics is approximately a major second' (p. 9).

Next we summarize a few studies that bear on the representation of pitch structures in two musical cultures, introducing each with a brief description of that culture's pitch system. In chapter 7, we also discussed related research on melodic expectations in Chinese and Sami Yoik music.

India: Rãg (or raga)

At the heart of Indian music is the rãg or raga, a sequence of notes taken from a scale (thãt), within which the performer improvises melodic and rhythmic structures. Each raga is a fixed set of notes associated with time of day, season of year, and expressive of some pattern of emotion and mood, according to Indian typology of emotions. There are ragas for morning and evening, for spring, summer, autumn, and winter, and so on. The traditional interpretation of the affective power of certain ragas is still widely accepted by Indian musicologists. The notes that constitute the 'scale' of a particular raga are drawn according to a precise set of principles. Two thãts, illustrated in Figure 15.2, are quite similar to the Western diatonic scale. However, there are eight other thãts, thus creating a wider range of scales than found in Western music. These scales create what has been called a 'circle of thãts.' The idea is very much like the circle of fifths in Western music, in that similarity relationships across thãts are best conceptualized as points around a circle.

One question that has been addressed in the research, then, is whether Indian and Western listeners perceive pitch relationships in rãgs consistent with the circle of thãts (Castellano, Bharucha, & Krumhansl, 1984). In Castellano et al.'s study, Indian and Western listeners were presented first with an Indian rãg, followed by one of several 'probe tones.' As in Krumhansl's other research (see chapter 5), listeners rated how well probe tones 'completed' the context. Indian listeners rated probe tones in ways that were consistent with Indian scale organization. Responses of Western listeners constituted a weaker approximation to the Indian scale structure. Interestingly, Western listeners did not assimilate pitches from Indian music to a 'Westernized' structure either. Rather Westerners appear to have formed a partial concept of Indian music structure through the context to which they were exposed during the experiment.

Indonesia: Pélog and sléndro

The Indonesian *gamelan* refers to an orchestra made up entirely of percussion instruments of various shapes, sizes, and materials. The metallic (usually bronze) instruments comprise gongs of several sizes of which the *gong besar* or large gong serves as the rhythmic 'engine,' and a collection of xylophone-like instruments usually provide the melodic components. In addition, there are drums, cymbals, bamboo rattles, and pipes, among other percussion instruments. Some of the instruments are shown in Figure 15.3, though the complete set is a larger collection than depicted in the figure. The *gamelan* has a revered place in Indonesian society. Author ST's father recalls that the *gamelan* belonging to the town where he lived was stored in a large covered courtyard in his family's home in Indonesia for many years (a responsibility they regarded as an honor), and the instruments were often retrieved for important festivities and town celebrations.

Gamelan tuning makes use of two systems that we may compare to 'scales,' though that may actually serve more as 'a set of guidelines for intervals' (Tenzer, 1991, p. 31). *Sléndro* is a five-tone system comprised of similar but somewhat unequal intervals (see Figure 15.2, third example under Indonesia). *Pélog* is a seven-tone system but commonly uses only five for a particular composition, depending on which 'mode' is in use; two variants are shown in Figure 15.2. Tenzer compares *sléndro* (very roughly) to the set of pitches A, C, D, E, G, transposed relative to the sequence shown in Figure 15.2, and *pélog* with the pitches E, F, G, A, B, C, D in its entirety (and E, F, G, B, C for one of its most common modes). However, this is only a rough comparison, as *gamelan* tones do not map exactly onto the tones of Western scales springing from the same tonic. For instance, there is no precise 2:1 frequency ratio between scale tones that are given the same name, and the distance between tones varies between instruments and changes with register (Tenzer, 2000, p. 30).

Empirical research suggests that listeners familiar with *gamelan* music appear to use pitch relationships found in associated scales when categorizing pitch intervals. In one study, Perlman and Krumhansl (1996) asked Indonesian and Western listeners to rate the size of melodic intervals. Certain participants showed subtle errors in the scaling of magnitude, and these confusions were predicted by the scale structure of their culture, which for Indonesian listeners included the *sléndro* and *pélog* scales (which, as can be seen in the examples above, are based on different intervals than diatonic scales). Thus although Western and Indonesian listeners differ in the kind of hierarchical (tonal) knowledge they use when listening to music, these listeners are alike in that they both use *some* kind of pitch hierarchy during listening. In chapter 8 we discussed how this internalization of a specific tonal

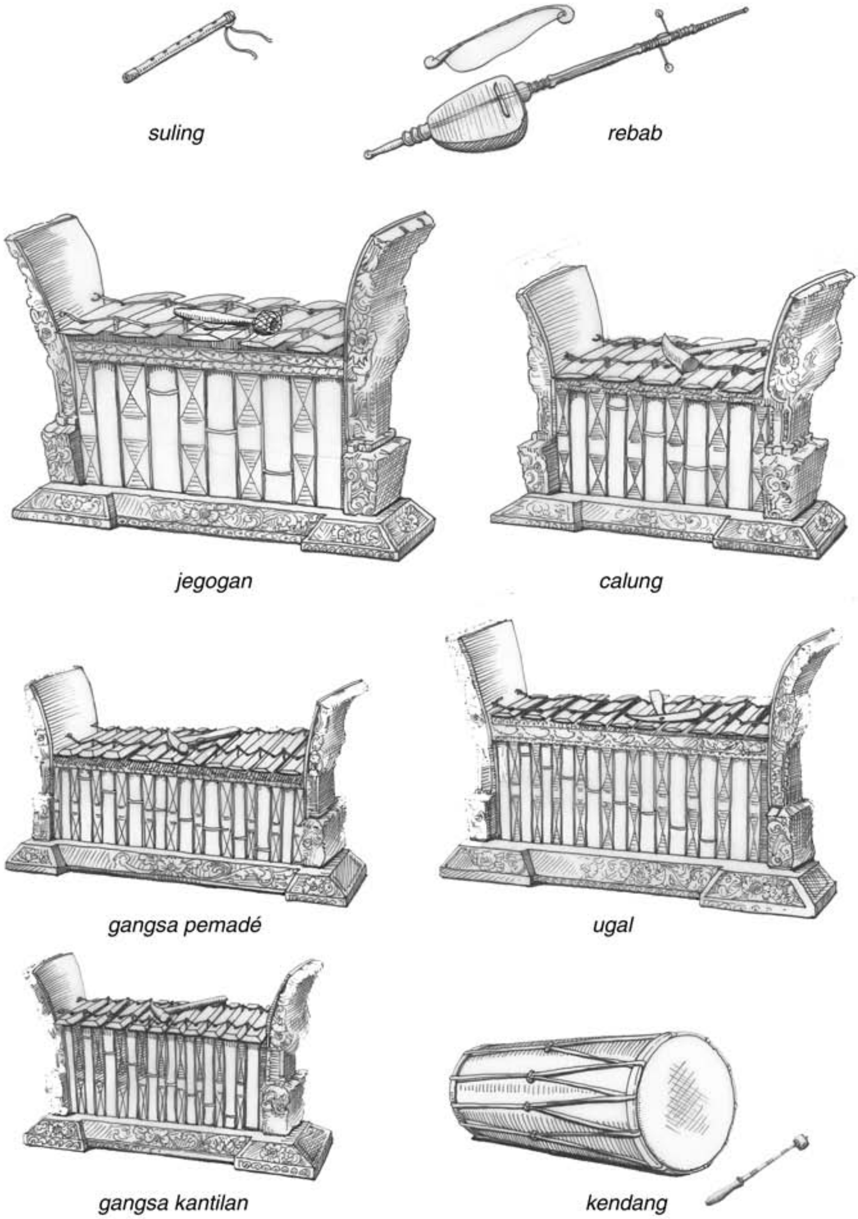
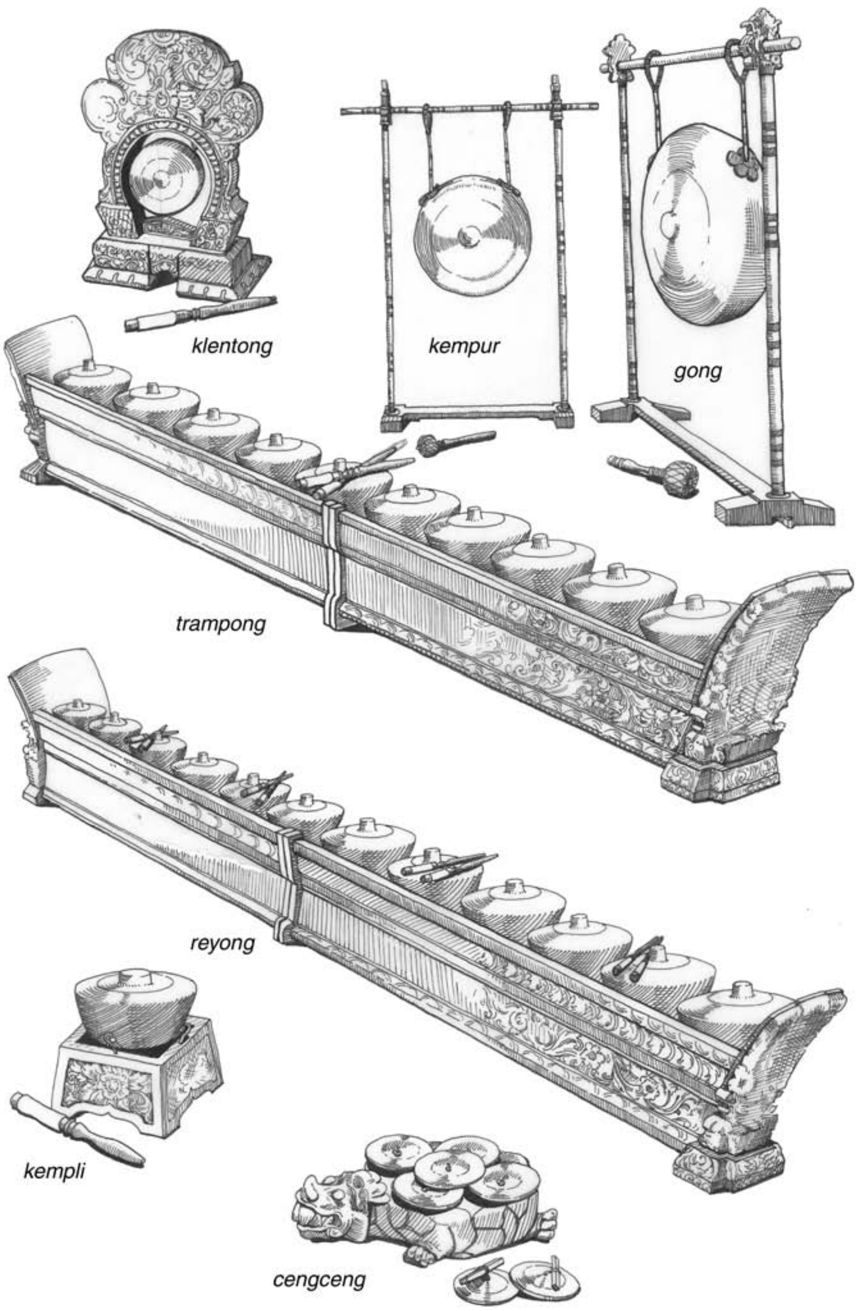


Figure 15.3 Artist's rendering of *gamelan gong kebyar* (Tenzer, 2000, pp. 42–43).

Source: Reprinted with permission from the University of Chicago Press.


klentong
kempur
gong
trampong
reyong
kempli
cengceng

scheme may occur fairly early in human development. For instance, Western infants discern mistuning of pitches equally well in melodies based on *pélog* or major/minor scales (Lynch et al., 1990). After about 1 year of age, however, they show greater sensitivity to discrepancies in tuning in melodies based on a major scale (Lynch & Eilers, 1992), performing similarly to Western adults.

Rhythm and metrical structure

Another fundamental way in which musical cultures differ is in rhythmic organization. One of the most common questions one hears when discussing rhythm concerns the rhythmic complexity of African drumming. Moreover, these drumming styles often do not suggest a simple metrical framework, often foiling the untutored listener (see chapter 6 for an example of a standard African drumming pattern that is difficult for most Westerners to reproduce). One main question regarding African drumming concerns whether rhythmic patterns conform to an overarching meter.

What makes African rhythms sound so different from Western rhythms? One view from ethnomusicology is that African rhythms are organized differently from Western rhythms. Specifically, African rhythms may not be metrical in the same sense as Western music. Whereas Western meters are based on the preservation of ratio relationships (see chapter 6), African rhythms are thought to be organized *additively*. Metrical organization implies a hierarchy, in that smaller time-spans are expressed as ratios of larger time-spans. By contrast, an additive organization (in this context) means that the organization is serial rather than hierarchical. Each time-span in an additive organization is autonomous, not linked to a larger superordinate span. In practice, the additive organization of African drumming involves the use of a fixed rhythmic pattern that functions like a 'timeline,' which provides the basis for all other instruments (Jones, 1959). Often the timeline is played by a bell. Though the timeline functions similarly to a 'beat,' in that it establishes a referent time series that instrumentalists use, it is usually structured from unequal time-spans. In a Western sense, then, the 'beat' of African drumming is syncopated rather than steady. This state of affairs is perplexing to the Westerner who is used to equating the beat with regularity (see chapter 6). Within this context, performers may syncopate with each other and to the timeline but the timeline itself – the foundation of timing – is thought to be structured additively.

Some empirical support for this view was reported by Magill and Pressing (1997). The authors recruited a master drummer who repeatedly produced a prototypical rhythmic pattern (shown in Figure 11.5). The authors used mathematical modeling (based on Wing & Kristofferson, 1973, discussed in chapter 11) to test hypotheses about the basis of the drummer's timing. One variant of this model incorporated a 'superordinate' timing structure that is consistent with the concept of meter. The authors found that this superordinate structure was not necessary: The drummer's timing patterns were consistent with a version in which all timed events were produced independently, as in the 'timeline' analogy described above.

However, the idea that African meter is ametrical has been questioned by Temperley (2000). Based primarily on analyses of transcriptions made by Blacking (1973) and A. M. Jones (1959), Temperley argues that there is pervasive evidence that African drumming is guided by metrical principles that are consistent with Lerdahl and Jackendoff's *Generative Theory of Tonal Music* (1983, see chapters 6 and 7). The striking differences between these musical traditions can be resolved by considering that syncopation plays a much larger role in African than in Western music (note that by invoking syncopation Temperley is also suggesting that a regular 'beat' underlies this music). African listeners are in a sense more 'comfortable' with syncopation than are Western listeners. In other words, what appears to be a difference in fundamental organization (metric versus additive) may instead be a difference in terms of how much rhythms are allowed to deviate from a single universal organizational scheme (meter).

Where do rhythmic differences come from? The answer to this question may remain largely unresolved, given the myriad historical and biological factors that influence a culture's music. Nevertheless, one potential source has been suggested recently: language. Patel and Daniele (2003) reported a provocative correlation between the rhythmic structure of language and the rhythmic structure found in instrumental music. They focused on a distinction among European languages. Some linguists refer to Germanic languages (which includes English and Dutch) as 'stress timed' and romantic languages (such as French and Spanish) as 'syllable timed' (e.g., Cruttenden, 1997). *Stress-timed* languages make distinctions between strong and weak syllables that can be measured in syllable duration and that relate to how vowels are produced. For instance the 'a' in 'about' is produced differently than the 'a' in 'tall' because it is linguistically reduced when used in the unstressed first syllable of 'about.' As a result, patterns of syllable timing in a language like English exhibit strong alternating contrasts between stressed and unstressed syllables. Interestingly, instrumental music by English-speaking classical composers (like Elgar) similarly exhibits strong contrasts in the duration of adjacent tones. For *syllable-timed* languages such as French, produced syllables are more equivalently timed. Consider the difference in the English pronunciation of 'chocolate,' with a lengthened first syllable, with the more regular timing of the three syllables in its French equivalent, 'chocolat'). French instrumental music (think Debussy) likewise exhibits greater regularity in the durations of adjacent tones than does English instrumental music. Patel and Daniele's findings point to the possibility that people are exposed to particular rhythmic patterns in language that are then carried over to the composition of music.

Related to the question of rhythm, there is also meter. As discussed in chapter 6, meters can vary from simple to complex, with complex meters involving cycles that do not reduce to simple binary or ternary patterns. In Western music these complex meters are rare, but they are common in other cultures. For instance, in South Indian classical music the *mridangam* (a two-sided pitched hand drum, featured in Figure 15.4) interweaves patterns on a wide variety of metrical cycles called *talas* that range from the very short (3 beats/cycle) to the very long (116 beats/cycle). *Talas* do not have fixed accenting patterns and this differentiates them from meters. Indian drumming features mathematical permutations and combinations in improvisatory embellishments as well as precomposed cadential figures that are conceived within various macro- and micro-subdivisions of the *tala* (Krishnamurthy, personal communication). In chapter 8 we discussed comparable research on the perception of complex meters that are common in Balkan music (Hannon & Trehub, 2005a). As noted in that chapter, the sense of meter can become deeply engrained in adulthood, although infant listeners appear to be equally able to learn rhythms that are presented within a simple and complex meter. From this perspective, it may be that children enter the world with universal predispositions for rhythm that over time develop into culturally specific tendencies.

Figure 15.4 Acclaimed Indian percussionist Rohan Krishnamurthy plays the *mridangam*, the principal percussion instrument of South India with a history of over 2000 years. The basic seating posture is said to be derived from yogic practices, and is aimed at creating a unison between player and instrument. The instrument is meant to be an extension of the player's body and not a separate entity.

Source: Photograph used with permission of Rohan Krishnamurthy and RohanRhythm. Copyright © RohanRhythm.

Phrase structure

As discussed in chapter 7, musical phrases – though salient to the listener – result from a complex combination of melodic, rhythmic, and harmonic factors. Research suggests that infants may be sensitive to musical phrasing from an early age (Jusczyk & Krumhansl, 1990). However, it is hard to test for subtleties of higher-order structure with infants, and other research suggests that certain aspects of phrase interpretation may be culturally specific. For instance, research comparing European with Arabic listeners suggests that both listeners were sensitive to rhythmic aspects of phrasing in Arabic instrumental improvisations but only Arabic listeners were sensitive to subtleties of the (modal) tonal system (Ayari & McAdams, 2003). Specifically, Arabic listeners heard segment boundaries when an instrumentalist switched from one mode to another within the Arabic modal system (maqām), as well as when a rhythmic group segmentation was conveyed (see chapter 7), whereas Western listeners – unfamiliar with Arabic modes – were only sensitive to boundaries established by rhythmic groups.

Another dimension is added when both language and music define the phrasing and melodic contour. Tonal languages such as Cantonese and Mandarin Chinese make use of melodic devices or tones as indicative of distinctions in meaning, whereas languages such as English use melodic devices mostly for paralinguistic distinctions, such as to convey a question versus a declaration. A comparative study of Cantonese and Mandarin music demonstrated a strong correspondence between lexical tones and melody in Cantonese music. Chan (1987) found that the lyrics of Cantonese songs were tonally fitted to the pitches of the melody, and lyric length to phrase length. In comparison, melody takes precedence in Mandarin songs and the original lexical tones of the lyrics were often ignored. The Cantonese language employs a more complex lexical tonal system, for instance employing more contrastive tones than Mandarin. Indeed, Kwan (1992) noted that matching Cantonese text to pre-existing popular melodies (such as in 'Cantopop') is a particularly difficult task that has led to a high-paying profession for lyricists in Hong Kong.

Performance

In the introduction to chapter 11 (see also chapter 12), we described a typical setting that one would experience at a Western classical music concert. As we noted, several aspects of that setting are specific to the particular musical culture. One aspect that is particularly important from a cultural perspective is the fact that we separated 'performer' from 'listener' at all! Western listeners tend to take this distinction for granted. In other cultures, however, such an idea is considered rather strange. In most of the world (and in Europe for thousands of years), music-making has been a collective activity, with all persons participating in both the production and perception of music. From this perspective, the fact that people are encouraged to stand and sing along at rock concerts should not be considered an oddity. Rather, it is the classical music lovers who engage in unusual behavior!

Another possible variation in the expression of music performance across cultures is that in music of many non-Western cultures, the integration of musical form with movement is closer than it is in Western music. For instance, John Baily (1985) analyzed the structure of Afghan music during a period in which a new stringed instrument, the Herati *dutãr* (a kind of lute), was developed. This 14-stringed instrument was conceptualized as a combination of the simple two-stringed instrument of the same name and another 14-stringed instrument, the *rubãb*. In comparing the original *dutãr* with the *rubãb*, Baily found that the structure of melodic movement and rhythmic organization for compositions from each instrument were constrained by the instrument design. For instance, music composed for the *rubãb* is characterized by a wider pitch range and more scalar movement (i.e., pitch motion in a continuing direction), presumably because of the larger number of strings in comparison with the 2-stringed original *dutãr*. Thus the range of motion a performer uses, and the correlated range of pitch, may be limited by the physical boundaries that constrain motion.

Related claims about African music have been made, with the additional suggestion that the experience of music in this culture has as much to do with the movements that create sound as with the sound itself (Kubik, 1962). This can best be understood, perhaps, when one experiences a live performance of an African drumming ensemble, which invariably includes dance as well as music. In this context, dance is an integral part of the musical experience, not an addition to it. In chapter 9, we also saw how movement and music are closely interwoven in young children's early musical explorations. Further, movement appears to be critical to the learning of rhythm and meter (chapter 8). Again the implication is that the Western 'art' music standard of the immobile audience (see chapters 11 and 12) may in fact be an odd way to experience music.

As compelling as these analyses of non-Western music are, one might wonder how different these analyses are from *all* Western music. As stated above, the sharpest distinctions arise when comparing the music of Afghanistan and African music with Western classical or 'art' music. In the case of Western popular and jazz music, on the other hand, it is common to associate movement – of the performers or of the audience members – with sound, thus suggesting greater commonality across these cultures. For instance Iyer (2002) has claimed that timing patterns characteristic of jazz (such as the delay of a snare drum that follows the bass drum) are best understood when one conceptualizes jazz as a way of simulating movement (in this case a hand clap that follows a foot stomp). Of course, it is arguably true that 'Western' forms such as rock music and jazz have more direct musicological connections to Africa than to Europe, and as such do not truly constitute 'Western' forms. However, it could also be argued that more 'Westerners' listen to these musical styles than to classical music!

Emotional interpretation

A primary concern for persons interested in cross-cultural differences related to music is the question of emotional communication. As discussed previously, if music is said to have 'meaning,' part of that meaning is the communication of emotion (chapter 13). One of the authors (PQP) recalls one of his first exposures to Chinese opera in Kaige Chen's film, *Farewell My Concubine* (1993). This film focuses on a specific sequence (from an opera of the same name) that is repeated several times. It sounded very strange to the author during the first presentation – interesting but not conveying any clear emotional message. By the end of the film, however, the music had become gradually more familiar. With that familiarity, the emotional contours of the sequence became clearer and the tragic message of the sequence came forth. It is plausible that such repeated exposure is necessary for the understanding of emotional communication from any 'foreign' culture.

Or is it? In fact, research suggests that listeners can at least identify the emotion being communicated in the music of a different culture even after the first exposure, even though the listener may not be 'moved' internally by that music as would a listener more familiar with that music. As discussed in chapter 14, Balkwill and Thompson (1999) found that listeners from North America could identify the intended emotions in Hindustani rãgs, based primarily on aspects of musical structure that are not specific to that cultural system such as tempo and pitch range. A similar implication comes out of an earlier study by Meyer, Palmer, and Mazo (1998) who found that Western listeners identified the emotional intention of Russian laments as more sorrowful and more internally coherent if recordings included a specific timbral cue. In this case, the timbral cue was a 'gasping' sound that occurs when the lamenter breathes in (an exaggerated and sustained bout of inspiration). Laments are characterized by long descending phrases in which the singer expresses her deep sorrow.

An important possibility to consider is that although the mode of musical expression varies greatly across cultures (e.g., tonal structure, rhythmic forms), there appears to be a universal set of emotional prototypes. This argument has been made strongly in a recent book entitled *The World in Six Songs* by Daniel Levitin (2008). According to Levitin, all cultures have produced types of songs that express important experiences held in common, which include friendship, joy, comfort, knowledge, religion, and love. By Levitin's account, music is an intrinsic part of human identity. Thus the messages communicated in music reflect core elements of humanity, and these elements come up as universal themes. Along similar lines, it has been observed by Sandra Trehub (2000) that music for children ubiquitously includes classes of lullabies and play songs, and that people can often recognize the distinction between these types of songs from another culture. Thus the close connection of music with one of the primary contexts of human life – the loving interactions between parent and child – begins right from the start.

Are there 'musical universals'?

It is common in summaries like this one to see authors offer a list of 'candidate' universal features (e. g., Carterette & Kendall, 1999; Justus & Hutsler, 2005; Stevens & Byron, 2009). Many have been mentioned in passing here; the list often includes items like the equivalence of pitches separated by the octave, organization of pitch into some kind of (typically asymmetric) tonal system, presence of the octave and perfect fifth in those tonal systems, the use of some kind of stable temporal referent (e.g., a pulse), and the segmentation of musical sequences into smaller groups or phrases. It is important to note the qualification here – nobody knows really which features are universal, particularly given the unevenness with which different cultures are represented. Nonetheless it is useful to consider whether a common 'core' of 'features' may exist.

The general consensus seems to be that universal features are probably general in scope, whereas cultures differ with respect to the details. For example, with respect to pitch, it has been proposed that a discrete pitch representation is universal, but the specific set of pitches varies. Similarly, asymmetric tonal scale systems may be universal though the specific set of asymmetric intervals in a scale may vary as we illustrated earlier. With respect to time, most seem to agree that rhythmic entrainment (see chapter 6) is probably universal, as is some kind of alternating pattern of emphasis, but the specific structures to which one entrains and the form of alternation may vary as in the West African 'timeline' discussed earlier.

Ultimately, of course, there is no easy 'yes' or 'no' answer to the question of universality. Music is intriguing in part because it is simultaneously culture-specific and universal. Music's meaning is simultaneously constrained by and extends across cultures. A comparison to language is informative. As with music, linguists typically believe that the world's languages comprise different specific manifestations of a common core of more general principles. However, music is unlike language in that the meaning of a foreign language is largely inaccessible to a nonspeaker. By contrast, research suggests that music can communicate its intention to the unschooled listener. At the same time, the intention behind music (as we discussed in chapter 13) is not fixed. Ultimately, it may be that music binds humanity together more effectively than does language, while at the same time acting as a vehicle for cultural diversity just like language does.

Final coda

Where does the existing research leave us? Obviously it is far too early to decide whether the significance of musical sound is a product of culturally learned norms or of universally shared neural and cognitive processes. It is of course likely that music results from interactions among culturally specific and universal components. The most interesting course for research to follow is to explore the expression of universal features in different cultures.

Although it may be premature to draw generalizations from the existing research, some intriguing hypotheses for future research can be derived. First, it appears as though cultures share similar 'core' features, but embellish these core features in ways that please different listeners. Thus cultural specificity may best be found in varieties of musical complexity. Second, some research cited above suggests that more 'universal' characteristics of music can be found in the time domain rather than in the pitch domain. Finally, with respect to performance it appears that there are distinctions between 'art' music and 'folk' (or 'popular') music that exist across cultures. Moreover, cross-cultural comparisons perhaps need to be careful not to conflate 'culture' with musical style (e.g., classical versus popular). These ideas are of course tentative and suggestive, not conclusive. As noted in our discussion, the cross-cultural focus is a 'growth' area in the psychology of music, and though it is developing rapidly, we have far to go.

We have intentionally used this chapter on culture and music as a concluding discussion to this book. Our perspective is that the psychology of music encompasses the scope of musical experience from the physical vibrations of sound to the kind of deep significance that leads people to invest so much time, money, and emotional energy in music and music-making. Moreover, the ubiquity of music across cultures – regardless of its commonalities and differences – argues for the fundamental importance of music to human life.

And so we close with a speculation about music's overall significance. Along the lines of Levitin's (2008) recent proposal (and echoing earlier Socratic thinking), it might be said that music functions as a mirror of human experience. The nature of this 'mirroring' was noted earlier (in chapter 13) in that music may be structurally 'isomorphic' to the internal experience of emotion (Langer, 1942). Moreover, to the degree that emotions are universal, certain elements of music may likewise be said to be universal. In this context, cultural differences may be considered to be analogous to differences across languages, used to communicate universal ideas and feelings. Just as languages communicate common ideas with different words and syntax, so does music communicate similar ideas through different tonal and rhythmic structures. The profound significance of music and the immense joy that it has given people in all times and places is almost certainly the result of the power of music to reveal the intricacies and depths of human life.

References

Abeles, H. F., Hoffer, C. R., & Klotman, R. H. (1994). *Foundations of music education* (2nd ed.). New York: Schirmer.

Abrams, R. M., Gerhardt, K., & Peters, A. J. M. (1995). Transmission of sound and vibration to the fetus. In J.-P. LeCanuet, W. Fifer, N. Krasnegor, & W. Smotherman (Eds.), *Fetal development: A psychobiological perspective* (pp. 315–330). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Abrams, R. M., Griffiths, K., Huang, X., Sain, J., Langford, G., & Gerhardt, K. J. (1998). Fetal music perception: The role of sound transmission. *Music Perception*, *15*, 307–317.

Adachi, M., & Trehub, S. E. (1998). Children's expression of emotion in song. *Psychology of Music*, *26*, 133–153.

Adamson, L. B., & Russell, C. L. (1999). Emotion regulation and the emergence of joint attention. In P. Rochat (Ed.), *Early social cognition: Understanding others in the first years of life* (pp. 281–297). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Aiello, R. (2001). Playing the piano by heart: From behavior to cognition. In R. J. Zatorre and I. Peretz (Eds.), *The Biological Foundations of Music. Annals of the New York Academy of Sciences, 930*, 389–393.

Allport, G. (1954). *The nature of prejudice*. Cambridge, MA: Addison-Wesley.

Altenmüller, E. O. (2003). How many music centers are in the brain? In I. Peretz & R. Zatorre (Eds.), *The cognitive neuroscience of music* (pp. 346–356). New York: Oxford University Press.

Amabile, T. M. (2001). Beyond talent: John Irving and the passionate craft of creativity. *American Psychologist*, *56*, 333–336.

American Heritage Dictionary of the English Language (2000, 4th ed.). New York: Houghton-Mifflin. [*Acoustics* appears on p. 15.]

Ando, Y. (1985). *Concert hall acoustics*. Berlin: Springer.

Ashley, R. (2009). *Musical improvisation*. In S. Hallam, I. Cross., & M. Thaut (Eds.), *The Oxford handbook of music psychology* (pp. 413–420). New York: Oxford University Press.

Athos, E. A., Levinson, B., Kistler, A., Zemansky, J., Bostrom, A., Freimer, N., & Gitschler, J. (2007). Dichotomy and perceptual distortions in absolute pitch ability. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 14795–14800.

Axelsson, A., & Lindgren, F. (1981). Hearing in classical musicians. *Acta Otolaryngologica Supplement*, *377*, 3–74.

Ayari, M., & McAdams, S. (2003). Aural analysis of Arabic improvised instrumental music (Taqsīm). *Music Perception*, *21*, 159–216.

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain*, *125*, 238–251.

Baddeley, A. D. (1986). *Working memory*. London: Oxford University Press.

Baily, J. (1985). Music structure and human movement. In I. Cross, R. West, & P. Howell (Eds.), *Musical structure and cognition* (pp. 237–258). London: Academic Press.

Balkwill, L.-L., & Thompson, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception*, *17*, 43–64.

Balkwill, L.-L., Thompson, W. F., & Matsunaga, R. (2004). Recognition of emotion in Japanese, Western, and Hindustani music by Japanese listeners. *Japanese Psychological Research*, *46*, 337–349.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Bandura, A., & Walters, R. (1963). *Social learning and personality development*. New York: Holt, Rinehart & Winston.

Bangert, M., & Schlaug, G. (2006). Specialization of the specialized in features of external human brain morphology. *European Journal of Neuroscience*, *24*, 1832–1834.

Bangerter, A., & Heath, C. (2004). The Mozart effect: Tracking the evolution of scientific legend. *British Journal of Psychology*, *43*, 605–623.

Bartlette, C., Headlam, D., Bocko, M., & Velikic, G. (2006). Effect of network latency on interactive musical performance. *Music Perception*, *24*, 49–62.

Beament, J. (1997). *The violin explained*. Oxford: Oxford University Press.

Beineix, J.-J. (Director), & Ossard, C. (Producer). (1981). *Diva* [Motion picture]. France: Les Films Galaxie. (Available from Anchor Bay Entertainment, 1699 Stutz Drive, Troy MI 48094.)

Benson, D. (2006). *Music: A mathematical offering*. Cambridge, UK: Cambridge University Press.

Beranek, L. L. (2004). *Concert halls and opera houses: Music, acoustics and architecture*. New York: Springer.

Beranek, L. L. (October, 2007). *Aspects of concert hall acoustics*. Heyser Memorial Lecture presented at the 123rd Convention of the Audio Engineering Society. Retrieved June 23, 2009, from http://www.aes.org/technical/heyser/downloads/AES123heyser-Beranek.pdf

Berger, A. A., & Cooper, S. (2003). Musical play: A case study of preschool children and parents. *Journal for Research in Music Education*, *51*, 151–165.

Bergeson, T. R., & Trehub, S. E. (2002). Absolute pitch and tempo in mothers' songs to infants. *Psychological Science*, *13*, 72–75.

Berkowitz, L. (1967). Experimental social psychology. In R. Harré (Ed.) *The sciences: Their origins and methods* (pp. 110–141). London: Blackie.

Bernstein, L. (1976). *The unanswered question: Six talks at Harvard*. Cambridge, MA: Harvard University Press.

Besson, M., & Schön, D. (2003). Comparison between language and music. In I. Peretz & R. Zatorre (Eds.), *The cognitive neuroscience of music* (pp. 269–293). New York: Oxford University Press.

Bever, T., & Chiarello, R. J. (1974, August 9). Cerebral dominance in musicians and nonmusicians. *Science*, *185*, 537–539.

Bharucha, J. J. (1984). Anchoring effects in music: The resolution of dissonance. *Cognitive Psychology*, *16*(4), 485–518.

Black, J. W. (1951). The effect of delayed side-tone upon vocal rate and intensity. *Journal of Speech and Hearing Disorders*, *16*, 56–60.

Blacking, J. (1973). *How musical is man?* Seattle, WA: University of Washington.

Blood, A. J., Zatorre, R. J., Bermudez, P., & Evans, A. C. (1999). Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions. *Nature Neuroscience*, *2*, 382–387.

Boerner, S. & von Streit, C. F. (2007). Promoting orchestral performance: The interplay between musicians' mood and a conductor's leadership style. *Psychology of Music*, *35*, 132–143.

Bolivar, V. J., Cohen, A. J., & Fentress, J. C. (1994). Semantic and formal congruency in music and motion pictures: Effects on the interpretation of visual action. *Psychomusicology*, *13*, 28–59.

Boltz, M. G. (2001). Musical soundtracks as a schematic influence on the cognitive processing of filmed events. *Music Perception*, *18*, 427–454.

Boltz, M. G. (2004). The cognitive processing of film and musical soundtracks. *Memory & Cognition*, *32*, 1194–1205.

Boltz, M., & Jones, M. R. (1986). Does rule recursion make melodies easier to reproduce? If not, what does? *Cognitive Psychology*, *18*, 389–431.

Boltz, M., Schulkind, M., & Kantra, S. (1991). Effects of background music on the remembering of filmed events. *Memory & Cognition*, *19*, 595–606.

Brackbill, Y., Adams, G., Crowell, D. H., & Gray, M. L. (1966). Arousal level in neonates and preschool children under continuous auditory stimulation. *Journal of Experimental Child Psychology*, *4*, 178–188.

Bradshaw, E., & McHenry, M. A. (2005). Pitch discrimination and pitch matching abilities of adults who sing inaccurately. *Journal of Voice*, *19*, 431–439.

Brattico, E., Tervaniemi, M., Näätänen, R., & Peretz, I. (2006). Musical scale properties are automatically processed in the human auditory cortex. *Brain Research*, *1117*, 162–174.

Bregman, A. S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.

Brochard, R., Abecasis, D., Potter, D., Ragot, R., & Drake, C. (2003). The 'ticktock' of our internal clock: Direct brain evidence of subjective accents in isochronous sequences. *Psychological Science*, *14*, 362–366.

Brooks, R. W. (1995). Mental practice and the musician: A practical approach to practice. *Update: Applications of Research in Music Education*, *13*, 4–8.

Brophy, T. S. (2005). A longitudinal study of selected characteristics of children's melodic improvisations. *Journal of Research in Music Education*, *53*, 120–133.

Brown, J. C. (2001). Feature dependence in the automatic identification of musical woodwind instruments. *Journal of the Acoustical Society of America*, *109*, 1064–1072.

Brown, S. (2000). The 'musilanguage' model of music evolution. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 271–300). Cambridge, MA: MIT Press.

Brown, S., Martinez, M. J., Hodges, D. A., Fox, P. T., & Parsons, L. M. (2004). The song system of the human brain. *Cognitive Brain Research*, *20*, 363–375.

Bruner, J. S. (1966). *Toward a theory of instruction*. Cambridge, MA: Harvard University Press.

Bullerjahn, C., & Güldenring, M. (1994). An empirical investigation of effects of film music using qualitative content analysis. *Psychomusicology*, *13*, 99–118.

Butler, C. (2004). *Pleasure and the arts*. Oxford: Oxford University Press.

Cahn, D. (2008). The effects of varying ratios of physical and mental practice, and task difficulty on performance of a tonal pattern. *Psychology of Music*, *36*, 179–191.

Caine, J. (1991). The effects of music on the selected stress behaviors, weight, caloric and formula intake, and length of hospital stay of premature and low birth weight neonates in NICU. *Journal of Music Therapy*, *28*, 180–192.

Cariani, P. A., & Delgutte, B. (1996). Neural correlates of the pitch of complex tones. Pitch and pitch salience. *Journal of Neurophysiology*, *76*, 1698–1716.

Carterette, E. C., & Kendall, R. A. (1999). Comparative music perception and cognition. In D. Deutsch (Ed.), *The psychology of music* (2nd ed., pp. 724–792). San Diego, CA: Academic Press.

Cassidy, J. W., & Standley, J. M. (1995). The effect of music listening on physiological responses of premature infants in the NICU. *Journal of Music Therapy. Special Issue: Special children's issue*, *32*, 208–227.

Castellano, M. A., Bharucha, J. J., & Krumhansl, C. L. (1984). Tonal hierarchies in the music of North India. *Journal of Experimental Psychology: General*, *113*, 394–412.

Castellano, G., Mortillaro, M., Camurri, A., Volpe, G., & Scherer, K. (2008). Automated analysis of body movement in emotionally expressive piano performances. *Music Perception*, *26*, 103–120.

Chabris, C. F. (1999). Prelude or requiem for the 'Mozart effect'? *Nature*, *400*, 826–827.

Chaffin, R., & Imreh, G. (2001). A comparison of practice and self-report as sources of information about the goals of expert practice. *Psychology of Music*, *29*, 39–69.

Chaffin, R., & Imreh, G. (2002). Practicing perfection: Piano performance as expert memory. *Psychological Science*, *13*, 342–349.

Chaffin, R., Imreh, G., & Crawford, M. (2002). *Practicing perfection: Memory and piano performance*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Chaffin, R., Lisboa, T., Logan, T., & Begosh, K. T. (2009, June 26). Preparing for memorized cello performance: the role of performance cues. *Psychology of Music*, DOI: 10.1177/0305735608100377.

Chan, M. K. M. (1987). Tone and melody interaction in Cantonese and Mandarin songs. *UCLA Working Papers in Phonetics*, *98*, 132–169.

Chang, H-.W., & Trehub, S. E. (1977). Auditory processing of relational information by young infants. *Journal of Experimental Child Psychology*, *24*, 324–331.

Chen, J., Woollacott, M. H., Pologe, S., & Moore, G. P. (2008). Pitch and space maps of skilled cellists: Accuracy, variability, and error correction. *Experimental Brain Research*, *188*(4), 493–503.

Chen, K. (Director), & Hsu, F. (Producer) (1993). *Farewell My Concubine* [Motion picture]. China: Miramax Films.

Chion, M. (1994). *Audio-vision: Sound on screen* (C. Gorbman, Ed. and Trans.). New York: Columbia University Press.

Chomsky, N. (1966). *Cartesian linguistics*. New York: Harper and Row.

Clarke E. F. (1987). Categorical rhythm perception: An ecological perspective. In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 19–33). Stockholm: Royal Swedish Academy of Music.

Clarke, E. [F.], & Baker-Short, C. (1987). The imitation of perceived rubato: A preliminary study. *Psychology of Music*, *15*, 58–75.

Coffman, D. D. (1990). Effects of mental practice, physical practice, and knowledge of results on piano performance. *Journal of Research in Music Education*, *38*, 187–196.

Cohen, A. J. (2001). Music as a source of emotion in film. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 249–272). Oxford: Oxford University Press.

Cohen, A. J. (2008). Music in performance arts: Film, theatre and dance. In S. Hallam, I. Cross, & M. Thaut (Eds), *The Oxford handbook of music psychology* (pp. 441–451). Oxford: Oxford University Press.

Colley, A. M., North, A. C., & Hargreaves, D. J. (2003). Gender bias in the evaluation of new age music. *Scandinavian Journal of Psychology*, *44*, 125–131.

Cook, N. (1987). The perception of large-scale tonal closure. *Music Perception*, *5*, 197–206.

Cooke, D. (1959). *The language of music*. Oxford: Oxford University Press.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, *61*, 1584–1594.

Costa-Giomi, E. (2004). Effects of three years of piano instruction on children's academic achievement, school performance, and self-esteem. *Psychology of Music*, *32*, 139–152.

Cottrell, N. B. (1972). Social facilitation. In C. G. McClintock (Ed.), *Experimental social psychology* (pp. 185–235). New York: Holt, Rinehart & Winston.

Creelman, C. D. (1962). Human discrimination of auditory duration. *Journal of the Acoustical Society of America*, *34*, 582–593.

Črnčec, R., Wilson, S. J., & Prior, M. (2006). No evidence for the Mozart effect in children. *Music Perception*, *23*, 305–317.

Cross, I. (2003). Music, cognition, culture, and evolution. In I. Peretz & R. Zatorre (Eds.), *The cognitive neuroscience of music* (pp. 42–56). New York: Oxford University Press.

Cross, I. (2005). Music and meaning, ambiguity, and evolution. In D. Miell, R. MacDonald, & D. J. Hargreaves (Eds.), *Musical communication* (pp. 27–43). Oxford: Oxford University Press.

Cross, I., & Tolbert, E. (2009). Music and meaning. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (pp. 24–34). Oxford: Oxford University Press.

Cruttenden, A. (1997). *Intonation* (2nd ed.). Cambridge, UK: Cambridge University Press.

Cuddy, L. L., Cohen, A. J., & Mewhort, D. J. K. (1981). Perception of structure in short melodic sequences. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 869–883.

Cuddy, L. L., & Lunney, C. A. (1995). Expectancies generated by melodic intervals: Perceptual judgments of melodic continuity. *Perception & Psychophysics*, *57*, 451–462.

Cuddy, L. L., Balkwill, L., Peretz, I., & Holden, R. R. (2005). Musical difficulties are rare: A study of 'tone deafness' among university students. *Annals of the New York Academy of Sciences*, *1060*, 311–324.

Custodero, L. A. (2002). The musical lives of young children: Inviting, seeking, and initiating. *Zero-to-Three Bulletin*, *23*, 4–9.

Custodero, L. A. (2006). Singing practices in 10 families with young children. *Journal of Research in Music Education*, *54*, 37–56.
Dahl, S., & Friberg, A. (2007). Visual perception of expressiveness in musicians' body movements. *Music Perception*, *24*, 433–455.
Dalla Bella, S., Giguère, J.-F., & Peretz, I. (2007). Singing proficiency in the general population. *Journal of the Acoustical Society of America*, *121*, 1182–1189.
Dalla Bella, S., Peretz, I., & Aronoff, N. (2003). The course of melody recognition: A gating paradigm study. *Perception & Psychophysics*, *65*, 1019–1028.
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.
Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: Grosset/Putnam.
Davidson, J. W. (1993). Visual perception of performance manner in the movements of solo musicians. *Psychology of Music*, *21*, 103–113.
Davidson, J. W. (1995). What does the visual information contained in music performances offer the observer? In R. Sternberg (Ed.), *The music machine: Psychophysiology and psychopathology of the sense of music*. Berlin: Springer-Verlag.
Davidson, J. W. (2005). Bodily communication in music performance. In D. Miell, R. MacDonald, & D. J. Hargreaves (Eds.), *Musical communication*. Oxford: Oxford University Press.
Davidson, J. W., & Good, J. M. (2002). Social and music co-ordination between members of a string quartet: An exploratory study. *Psychology of Music*, *30*, 186–201.
Davidson, L. (1994). Songsinging by young and old: A developmental approach to music. In R. Aiello & J. A. Sloboda (Eds.), *Musical perceptions* (pp. 99–130). New York: Oxford University Press.
Davidson, L., McKernon, P., & Gardner, H. (1981). The acquisition of song: A developmental approach. In J. A. Mason (Ed.), *Documentary report of the Ann Arbor symposium on the application of psychology to the teaching and learning of music* (pp. 301–315). Reston, VA: Music Education National Conference.
Davies, C. (1994). The listening teacher. An approach to the collection and study of invented songs of children aged 5 to 7. In H. Lees (Ed.), *Musical connections: Tradition and change* (pp. 120–127). Auckland, NZ: International Society for Music Education.
Davies, J. B. (1978). *The psychology of music*. Stanford, CA: Stanford University Press.
Davis, S. (2006). Implied polyphony in the solo string works of J. S. Bach: A case for the perceptual relevance of structural expression. *Music Perception*, *23*, 423–446.
Deary, I. J., & Matthews, G. (1993). Personality traits are alive and well. *The Psychologist*, *6*, 299–311.
DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, *208*, 1174–1176.
DeCasper, A. J., & Prescott, P. A. (1984). Human newborns' perception of male voices: Preference, discrimination, and reinforcing value. *Developmental Psychobiology*, *17*, 481–491.
DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior & Development*, *9*, 133–150.
De L'Etoile, S. K. (2006). Infant-directed singing: A theory for clinical intervention. *Music Therapy Perspectives*, *24*, 22–29.

Deliège, I. (1987). Grouping conditions in listening to music: An approach to Lerdahl & Jackendoff's grouping preference rules. *Music Perception*, *4*, 325–360.

DeNora, T. (2001). Aesthetic agency and musical practice: New directions in the sociology of music and emotion. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 161–180). New York: Oxford University Press.

Desain, P., & Honing, H. (2003). The formation of rhythmic categories and metric priming. *Perception*, *32*, 341–365.

Deutsch, D. (1972, March 3). Mapping of interactions in the pitch memory store. *Science*, *175*, 1020–1022.

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

Deutsch, D. (1995). *Musical illusions and paradoxes* [CD]. La Jolla, CA: Philomel Records.

Deutsch, D. (1999). Grouping mechanisms in music. In D. Deutsch (Ed.), *The psychology of music* (2nd ed., pp. 299–348). San Diego, CA: Academic Press.

DeWitt, L. A., & Samuel, A. G. (1990). The role of knowledge-based expectations in music perception: Evidence from musical restoration. *Journal of Experimental Psychology: General*, *119*, 123–144.

Dickey, M. R. (1991). A comparison of verbal instruction and nonverbal teacher-student modeling in instrumental ensembles. *Journal of Research in Music Education*, *39*, 132–142.

Dissanayake, E. (2000). Antecedents of the temporal arts in early mother-infant interaction. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 389–410). Cambridge, MA: MIT Press.

Doane, M. A. (1980). The voice in the cinema: The articulation of body and space. *Yale French Studies*, *60*, p. 34.

Dowling, W. J. (1968). *Rhythmic fission and the perceptual organization of tone sequences*. Unpublished doctoral dissertation, Harvard University.

Dowling, W. J. (1973). The perception of interleaved melodies. *Cognitive Psychology*, *5*, 322–337.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, *85*, 341–354.

Dowling, W. J. (1984). Development of musical schemata in children's spontaneous singing. In W. R. Crozier & A. J. Chapman (Eds.), *Cognitive processes in the perception of art* (pp. 145–164). Amsterdam: North-Holland.

Dowling, W. J., & Bartlett, J. C. (1981). The importance of interval information in long-term memory for melodies. *Psychomusicology*, *1*, 30–49.

Dowling, W. J., Kwak, S., & Andrews, M. W. (1995). The time course of recognition of novel melodies. *Perception & Psychophysics*, *57*, 136–149.

Drake, C., & Botte, M.-C. (1993). Tempo sensitivity in auditory sequences: Evidence for a multiple-look model. *Perception & Psychophysics*, *54*, 277–286.

Drake, C., Jones, M. R., & Baruch, C. (2000). The development of rhythmic attending in auditory sequences: Attunement, referent period, focal attending. *Cognition*, *77*, 251–288.

Drake, C., Penel, A., & Bigand, E. (2000). Tapping in time with mechanically and expressively expressively performed music. *Music Perception*, *18*, 1–25.

Eaton, S., & Gillis, H. (2002). Review of orchestra musicians' hearing loss risks. *Canadian Acoustics*, *30*, 5.

Eitan, Z., & Granot, R. Y. (2008). Growing oranges on Mozart's apple tree: 'Inner form' and aesthetic judgment. *Music Perception*, *25*, 397–418.

Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B., & Taub, E. (1995, October 13). Increased cortical representation of the fingers of the left hand in string players. *Science*, *270*, 305–307.

Epstein, D. (1995). *Shaping time: Music, the brain, and performance*. New York: Schirmer.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*, 363–406.

Farnsworth, P. R. (1958). *The social psychology of music*. Austin, TX: Holt, Rinehart, & Winston.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development*, *60*, 1497–1510.

Fernald, A. (2001). Hearing, listening, and understanding: Auditory development in infancy. In G. Bremner & A. Fogel (Eds.), *Blackwell handbook of infant development* (pp. 35–70). Oxford: Blackwell.

Ferrier, L. (1985). Intonation and discourse: Talk between 12-month-olds and their mothers. In K. E. Nelson (Ed.), *Children's language* (Vol. 5, pp. 35–60). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Field, T. (1998). Maternal depression effects on infants and early interventions. *Preventive Medicine*, *27*, 2000–2003.

Fifer, W. P., & Moon, C. (1989). Psychobiology of newborn auditory preferences. *Seminars in Perinatology*, *13*, 430–433.

Fine, P., Berry, A., & Rosner, B. (2006). The effect of pattern recognition and tonal predictability on sight-singing ability. *Psychology of Music*, *34*, 431–447.

Finney, S. A. (1997). Auditory feedback and musical keyboard performance. *Music Perception*, *15*, 153–174.

Fischer, S., Hallschmid, M., Elsner, A. L., & Born, J. (2002). Sleep forms memory for finger skills. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 11987–11991.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.

Fogel, H. (2007). Formality in concert: Have we gone too far? Blog entry December 12, 2007 and December 11, 2007. Retrieved June 20, 2009, from http://www.artsjournal.com/ontherecord

Ford, L., & Davidson, J. W. (2003). An investigation of members' roles in wind quintets. *Psychology of Music*, *31*, 53–74.

Forsythe, M. (1985). *Buildings for music*. Cambridge, MA: MIT Press.

Fox, D. B. (1989). MusicTIME and music times two. In B. Andress (Ed.), *Prekindergarten music education* (pp. 13–24). Reston, VA: Music Educators National Conference.

Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *Psychology of music* (pp. 149–181). San Diego, CA: Academic Press.

Francès, R. (1958/1988). *The perception of music*. (W. J. Dowling, Trans.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Frankland, B. W., & Cohen, A. J. (2004). Parsing of melody: Quantification and testing of the local grouping rules of Lerdahl and Jackendoff's A Generative Theory of Tonal Music. *Music Perception*, *21*, 499–543.

Freud, S. (1924). The passing of the Oedipus-complex. In *Collected papers*, *Vol. II*. London: Hogarth.

Friberg, A., & Sundberg, J. (1999). Does music performance allude to locomotion? A model of final *ritardandi*, derived from measurements of stopping runners. *Journal of the Acoustical Society of America*, *105*, 1469–1484.

Fujioka, T., Ross, B., Kakigi, R., Pantev, C., & Trainor, L. (2006). One year of musical training affects development of auditory cortical-evoked fields in young children. *Brain*, *129*, 2593–2608.

Gabriel, C. (1978). An experimental study of Deryck Cooke's theory of music and meaning. *Psychology of Music*, *6*, 13–20.

Gabrielsson, A. (1999). The performance of music. In D. Deutsch (Ed.), *The psychology of music* (pp. 501–602). San Diego, CA: Academic Press.

Gabrielsson, A. (2001). Emotions in strong experiences with music. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 431–449). Oxford: Oxford University Press.

Gabrielsson, A. (2003). Music performance research at the millennium. *Psychology of Music*, *31*, 221–272.

Gabrielsson, A., & Lindstrom, E. (2001). The influence of musical structure on emotional expression. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 223–248). Oxford: Oxford University Press.

Gade, A. C. (2007). Acoustics in halls for speech and music. In T. D. Rossing (Ed.), *Springer handbook of acoustics* (pp. 301–350). New York: Springer.

Galton, F., Sir (1979). *Hereditary genius: An inquiry into its laws and consequences*. London: Julian Friedman. (Original work published 1869)

Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *Journal of Neuroscience*, *23*, 9240–9245.

Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. (2002). *Cognitive neuroscience: The biology of the mind* (2nd ed.). New York: W. W. Norton.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton-Mifflin.

Gieseking, W., & Leimer, K. (1972). *Piano technique: The shortest way to pianistic perfection*. New York: Dover. (Original work published 1932)

Ginsborg, J. (2002). Classical singers learning and memorizing a new song: An observational study. *Psychology of Music*, *30*, 58–101.

Ginsborg, J., & Sloboda, J. A. (2007). Singers' recall for the words and melody of a new, unaccompanied song. *Psychology of Music*, *35*, 421–440.

Givens, B. (1996). Stimulus-evoked resetting of the dentate theta rhythm: Relation to working memory. *NeuroReport*, *8*, 159–163.

Goebl, W. (2001). Melody lead in piano performance: Expressive device or artifact? *Journal of the Acoustical Society of America*, *110*, 563–572.

Goebl, W., & Palmer, C. (2009). Synchronization of timing and motion among performing musicians. *Music Perception*, *26*, 427–438.

Gorali-Turel, T. (1999). Spontaneous movement response of young children to musical stimulation as indicator of the hidden cognitive process. Unpublished paper. *Cognitive processes of children engaged in musical activity*, Urbana-Champaign, IL: School of Music, University of Illinois. [Cited in McPherson, G. (2006). *The child as musician* (p. 296). Oxford: Oxford University Press.]

Gorbman, C. (1987). *Unheard melodies: Narrative film music*. Bloomington, IN: Indiana University Press.

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.

Green, G. A. (1990). The effect of vocal modeling on pitch-matching accuracy of elementary schoolchildren. *Journal of Research in Music Education*, *38*, 225–331.

Greengrass, P. (Director), & Powell, J. (Original score). (2004). *The Bourne supremacy* [Motion picture]. United States/Germany: Universal Pictures.

Greengrass, P. (Director), & Powell, J. (Original score). (2007). *The Bourne ultimatum* [Motion picture]. United States/Germany: Universal Pictures.

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

Gruson, L. M. (1988). Rehearsal skill and musical competence: Does practice make perfect? In J. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation, and composition* (pp. 91–112). Oxford: Oxford University Press.

Guhn, M., Hamm, A., & Zentner, M. (2007). Physiological and musico-acoustic correlates of the chill response. *Music Perception*, *24*, 473–483.

Gurney, N., & Gurney, E. (1965). *The king, the mice, and the cheese*. New York: Beginner Books/Random House.

Hajda, J. M., Kendall, R. A., Carterette, E. C., & Harshberger, M. L. (1997). Methodological issues in timbre research. In I. Deliège & J. Sloboda (Eds.), *Perception and cognition of music* (pp. 253–306). Hove, UK: Psychology Press.

Hall, D. E. (2001). *Musical acoustics*. Pacific Grove, CA: Brooks/Cole.

Hallam, S. (1995). Professional musicians' approaches to the learning and interpretation of music. *Psychology of Music*, *23*, 111–128.

Hallam, S. (1997). Approaches to instrumental music practice of experts and novices: Implications for education. In H. Jørgensen & A. C. Lehmann (Eds.), *Does practice make perfect? Current theory and research on instrumental music practice* (pp. 89–107). Oslo, Norway: Norges Musikkhøgskole.

Halpern, A. (2003). Cerebral substrates of musical imagery. In I. Peretz & R. Zatorre (Eds.), *The cognitive neuroscience of music* (pp. 217–230). New York: Oxford University Press.

Hamilton, K. (2007). *After the Golden Age: Romantic pianism and modern performance*. Oxford: Oxford University Press.

Handel, S. (1989). *Listening: An introduction to the perception of auditory events*. Cambridge, MA: MIT Press.

Hannon, E. E., & Trehub, S. E. (2005a). Metrical categories in infancy and adulthood. *Psychological Science*, *16*, 48–55.

Hannon, E. E., & Trehub, S. E. (2005b). Tuning in to musical rhythms: Infants learn more readily than adults. *Proceedings of the National Academy of Sciences*, *102*, 12639–12643.

Hanslick, E. (1986). *On the musically beautiful*. (G. Payzant, Trans.). Indianapolis, IN: Hackett. (Original work published 1891)

Hargreaves, D. J. (1986). *The developmental psychology of music*. Cambridge, UK: Cambridge University Press.

Hargreaves, D. J., & North, A. C. (Eds.) (1997). *The social psychology of music*. Oxford: Oxford University Press.

Harré, R. (1997). Emotion in music. In M. Hjort & S. Laver (Eds.), *Music and the arts* (pp. 110–118). New York & Oxford: Oxford University Press.

Harré, R., & Parrott, W. G. (1996). *The emotions: Social, cultural and biological dimensions*. London and Thousand Oaks, CA: Sage.

Harré, R., & Secord, P. F. (1973). *The explanation of social behaviour*. Oxford: Blackwell.

Harrison, A. C., & O'Neill, S. A. (2000). Children's gender-typed preferences for musical instruments: An intervention study. *Psychology of Music*, *28*, 81–97.

Hayes, J. R. (1981). *The complete problem solver*. Philadelphia, PA: Franklin Institute Press.

He, C., Hotson, L., & Trainor. L. J. (2007). Mismatch responses to pitch changes in early infancy. *Journal of Cognitive Neuroscience*, *19*, 878–892.

He, C., Hotson, L., & Trainor, L. J. (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.

Helmholtz, H. L. F. von. (1863). *On the sensations of tone as the physiological basis of the theory of music*. New York: Dover.

Henley, P. T. (2001). Effects of modeling and tempo patterns as practice techniques on the performance of high school instrumentalists. *Journal of Research in Music Education*, *49*, 169–180.

Hepper, P. G. (1991). An examination of fetal learning before and after birth. *The Irish Journal of Psychology*, *12*, 95–107.

Hevner, K. (1936). Experimental studies of the elements of expression in music. *American Journal of Psychology*, *48*, 246–268.

Hicock, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory-motor interaction revealed by fMRI: Speech, music, and working memory in area SPT. *Journal of Cognitive Neuroscience*, *15*, 673–682.

Highben, Z., & Palmer, C. (2004). Effects of auditory and motor mental practice in memorized piano performance. *Bulletin of the Council for Research in Music Education*, *159*, 58–65.

Hirsch, I. J. (1959). Auditory perception of temporal order. *Journal of the Acoustical Society of America*, *31*, 759–767.

Hitchcock, A. (Director), & Herrmann, B. (Original score). (1960). *Psycho* [Motion picture]. United States: Shamley Productions.

Hofman, P. M., Van Riswick, J. G. A., & Van Opstal, A. J. (1998). Relearning sound localization with new ears. *Nature Neuroscience*, *1*, 417–421.

Honing, H. (2005). Is there a perception-based alternative to kinematic models of tempo rubato? *Music Perception*, *23*, 79–85.

Hood, M. (1966). *Sléndro* and *pélog* redefined. *Selected Reports*, UCLA Institute of Ethnomusicology, 1, 28–37.

Hospers. J. (1964). *Meaning and truth in the arts*. Hamden, CN: Archon Books.

Howe, M. J. A., Davidson, J. W., & Sloboda, J. A. (1998). Innate talents: Reality or myth? *Behavioral and Brain Sciences*, *21*, 399–442.

Howell, P., Powell, D. J., & Khan, I. (1983). Amplitude contour of the delayed signal and interference in delayed auditory feedback tasks. *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 772–784.

Hulse, S. H., Cynx, J., & Humpal, J. (1984). Absolute and relative pitch discrimination in serial pitch perception by birds. *Journal of Experimental Psychology: General*, *113*, 38–54.

Huron, D. (2003). Is music an evolutionary adaptation? In I. Peretz & R. Zatorre (Eds.), *The cognitive neuroscience of music* (pp. 57–78). New York: Oxford University Press.

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. Cambridge, MA: MIT Press.

Hutchins, S., & Palmer, C. (2008). Repetition priming in music. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 693–707.

Hyde, K. L., Lerch, J. P., Zatorre, R. J., Griffiths, T. D., Evans, A. C., & Peretz, I. (2007). Cortical thickness in congenital amusia: When less is better than more. *Journal of Neuroscience*, *27*, 13028–13032.

Hyde, K. L., Zatorre, R. J., Griffiths, T. D., Lerch, J. P., & Peretz, I. (2006). Morphometry of the amusic brain: A two-site study. *Brain*, *129*, 2562–2570.

Ilari, B., & Polka, L. (2006). Music cognition in early infancy: Infants' preferences and long-term memory for Ravel. *International Journal of Music Education*, *24*, 7–20.

Ingram, J. (1997, May 11). More proof music lifts young IQs. *The Toronto Star*, p. F8.

Ivanov, V. K., & Geake, J. G. (2003). The Mozart Effect and primary school children. *Psychology of Music*, *31*, 405–413.

Ivry, R. B., & Spencer, R. M. C. (2004). The neural representation of time. *Current Opinion in Neurobiology*, *14*, 225–232.

Iyer, V. (2002). Embodied mind, situated cognition, and expressive microtiming in African-American music. *Music Perception*, *19*, 387–414.

James, W. (1890). *Principles of psychology* (Vol. 2). New York: Holt.

Janata, P., Birk, J., van Horn, J., Leman, M., Tillmann, B., & Bharucha, J. (2002). The cortical topography of tonal structures underlying Western music. *Science*, *298*, 2167–2170.

Jarvis, E. D. (2004). Learned birdsong and the neurobiology of human language. *Annals of the New York Academy of Sciences*, *1016*, 746–777.

Johnson-Laird, P. N. (2002). How jazz musicians improvise. *Music Perception*, *19*, 415–442.

Jones, A. M. (1959). *Studies in African music*. London: Oxford.

Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review*, *83*, 323–355.

Jones, M. R. (1987). Dynamic pattern structure in music: Recent theory and research. *Perception & Psychophysics*, *41*, 621–634.

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, *96*, 459–491.

Jones, M. R., Moynihan, H. M., MacKenzie, N., & Puente, J. K. (2002). Stimulus-driven attending in dynamic auditory arrays. *Psychological Science*, *13*, 313–319.

Jusczyk, P. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Jusczyk, P., & Krumhansl, C. L. (1990). Infants' perception of phrase structure in music. *Psychological Science*, *1*, 70–73.

Juslin, P. N. (1997). Emotion communicated in music performance: A functionalist perspective and some data. *Music Perception*, *14*, 383–418.

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1797–1813.

Juslin, P. N. (2001). Communicating emotion in performance: A review and a theoretical approach. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 309–337). Oxford: Oxford University Press.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*, 770–814.

Juslin, P. N., & Sloboda, J. A. (2001). *Music and emotion: Theory and research*. Oxford: Oxford University Press.

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, *31*, 559–621.

Justus, T., & Hutsler, J. J. (2005). Fundamental issues in the evolutionary psychology of music: Assessing innateness and domain specificity. *Music Perception*, *23*, 1–28.

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

Kammler, D. (2008). *A first course in Fourier analysis*. Cambridge, UK: Cambridge University Press.

Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (1995). *Essentials of neural science and behavior*. Norwalk, CT: Appleton and Lange.

Keller, P. E. (2001). Attentional resource allocation in musical ensemble performance. *Psychology of Music*, *29*, 20–38.

Kemp, A. E. (1996). *The musical temperament: Psychology and personality of musicians*. Oxford: Oxford University Press.

Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, *11*, 201–258.

Kendall, L. (1993). Liner notes analysis accompanying *Original soundtrack anthology: Star Wars trilogy*. New York, NY: Lucasfilm/Arista/Twentieth Century Fox.

Kendall, R. A., (1986). The role of acoustic signal partitions in listener categorization of musical phrases. *Music Perception, 4*, 185–214.

Kessen, W., Levine, J., & Wendrich, K. (1979). The imitation of pitch in infants. *Infant Behavior & Development*, *2*, 93–99.

Kieslowski, K. (Director), & Karmitz, M. (Producer). (1993). *Three colors: Blue* [Motion picture]. France: CAB Productions. (Available from Miramax Films)

Kisilevsky, B. S., & Low, J. A. (1998). Human fetal behavior: 100 years of study. *Developmental Review*, *18*, 1–29.

Kisilevsky, B. S., Hains, S. M. J., Jacquet, A.-Y., Granier-Deferre, C., & Lecanuet, J. P. (2004). Maturation of fetal responses to music. *Developmental Science*, *7*, 550–559.

Kisilevsky, B. S., Killen, H., Muir, D. W., & Low, J. A. (1991). Maternal and ultrasound measurements of elicited fetal movements. *Obstetrics and Gynecology*, *77*, 889–892.

Kivy, P. (1990). *Music alone*. Cambridge, MA: Harvard University Press.

Knudsen, E. I., & Knudsen, P. F. (1989). Vision calibrates sound localization in developing barn owl. *Journal of Neuroscience*, *9*, 3306–3313.

Koelsch, S., Gunter, T. C., v. Cramon, D. Y., Zysset, S., Lohmann, G., & Frederici, A. D. (2002). Bach speaks: A cortical 'language-network' serves the processing of music. *NeuroImage*, *17*, 956–966.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Frederici, A. D. (2004). Music, language and meaning: Brain signatures of semantic processing. *Nature Neuroscience*, *7*, 302–307.

Koffka, K. (1935). *Principles of Gestalt psychology*. New York: Harcourt Brace.

Konečni, V. (2008). Does music induce emotion? A theoretical and methodological analysis. *Psychology of Aesthetics, Creativity and the Arts*, *2*, 115–129.
Kopiez, R., & Lee, J. I. (2006). Towards a dynamic model of skills involved in sight reading music. *Music Education Research*, *8*, 97–120.
Koskoff, E. (1995). When women play: The relationship between musical instruments and gender style. *Canadian University Music Review*, *16*, 114–127.
Krampe, R. T., & Ericsson, K. A. (1995). Deliberate practice and elite musical performance. In J. Rink (Ed.), *The practice of performance: Studies in musical interpretation* (pp. 84–102). Cambridge, UK: Cambridge University Press.
Kratus, J. (1989). A time analysis of the compositional processes used by children ages 7 to 11. *Journal of Research in Music Education*, *37*, 5–20.
Krueger, C., Holditch-Davis, D., Quint, S., & DeCasper, A. (2004). Recurring auditory experience in the 28- to 34-week-old fetus. *Infant Behavior & Development*, *27*, 537–543.
Krumhansl, C. L. (1979). The psychological representation of musical pitch in a tonal context. *Cognitive Psychology*, *11*, 346–374.
Krumhansl, C. (1990). *Cognitive foundations of musical pitch*. Oxford: Oxford University Press.
Krumhansl, C. L. (1995). Music psychology and music theory: Problems and prospects. *Music Theory Spectrum*, *17*, 53–80.
Krumhansl, C. L. (1996). A perceptual analysis of Mozart's Piano Sonata K. 282: Segmentation, tension, and musical ideas. *Music Perception*, *13*, 401–432.
Krumhansl, C. L. (1997). An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology*, *51*, 336–353.
Krumhansl, C. L., 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.
Kubik, G. (1962). The phenomenon of inherent rhythms in East and Central African instrumental music. *African Music*, *3*, 33–42.
Kwan, K. (1992). Textual and melodic contour in Cantonese popular songs. *Secondo Convegno Europeo di Analisi Musicale* [Symposium Proceedings for the Second European Meeting of Musical Analysis]. Trento, Italy: Università degli Studi.
Lahav, A., Saltzman, E., & Schlaug, G. (2007). Action representation of sound: Audiomotor recognition network while listening to newly acquired actions. *Journal of Neuroscience*, *27*, 308–314.
Lamont, A. (1998). Music, education, and the development of pitch perception: The role of context, age, and musical experience. *Psychology of Music*, *26*, 7–25.
Langer, S. K. (1942). *Philosophy in a new key*. Cambridge, MA: Harvard University Press.
Langer, S. K. (1953). *Feeling and form*. London: Routledge and Kegan Paul.
Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, *106*, 119–159.
Large, E. W., Palmer, C., & Pollack, J. B. (1995). Reduced memory representations for music. *Cognitive Science*, *19*, 53–96.
Lashley, K. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–136). New York: Wiley.
Laukka, P., & Gabrielsson, A. (2000). Emotional expression in drumming performance. *Psychology of Music*, *28*, 181–189.

Le Bon, G. (1947). *The crowd: A study of the popular mind*. London: Benn. (Original work published 1895)

Lecanuet, J.-P., & Granier-Deferre, C. (1993). Speech stimuli in the fetal environment. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition* (pp. 237–248). New York: Plenum.

Lee, B. S. (1950). Effects of delayed speech feedback. *Journal of the Acoustical Society of America*, *22*, 824–826.

Lehmann, A. C., & Ericsson, K. A. (1997). Research on expert performance and deliberate practice: Implications for the education of amateur musicians and music students. *Psychomusicology*, *16*, 40–58.

Lempert, T., & Bauer, L. T. (1995). Mass fainting at rock concerts. *New England Journal of Medicine*, *332*, 1721.

LePage, E. L., & Murray, N. M. (1998). Latent cochlear damage in personal stereo users: A study based on click-evoked otoacoustic emissions. *Medical Journal of Australia*, *169*, 588–592.

Lerdahl, F. (1988). Cognitive constraints on compositional systems. In J. A. Sloboda (Ed.), *Generative processes in music* (pp. 231–259). Oxford: Clarendon Press.

Lerdahl, F. (2001). *Tonal pitch space*. New York: Oxford University Press.

Lerdahl, F. (2009). Genesis and architecture of the GTTM project. *Music Perception*, *26*, 187–194.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Lerdahl, F., & Jackendoff, R. (1983–84). An overview of hierarchical structure in music. *Music Perception*, *1*, 229–252.

Lerdahl, F., & Krumhansl, C. L. (2007). Modeling tonal tension. *Music Perception*, *24*, 329–366.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levinson, J. (1997). *Music in the moment*. Ithaca, NY: Cornell University Press.

Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception & Psychophysics*, *56*, 414–423.

Levitin, D. J. (2006). *This is your brain on music: The science of a human obsession*. New York: Penguin.

Levitin, D. J. (2008). *The world in six songs: How the musical brain created human nature*. New York: Dutton.

Levitin, D. J., & Menon, V. (2003). Musical structure is processed in 'language' areas of the brain: A possible role for Brodmann Area 47 in temporal coherence. *NeuroImage*, *20*, 2142–2152.

Lexmann, J. (2006). *Theory of film music*. Frankfurt: Peter Lang.

Liégois-Chauvel, C., Peretz, I., Babaï, M., Laguitton, V., & Chauvel, P. (1998). Contribution of different cortical areas in the temporal lobes to music processing. *Brain*, *121*, 1853–1867.

Lim, S., & Lippman, L. G. (1991). Mental practice and memorization of piano music. *Journal of General Psychology*, *118*, 21–30.

Liman, D. (Director), & Powell, J. (Original score). (2002). *The Bourne identity*. [Motion picture]. United States/Germany/Czech Republic: Universal Pictures.

Limb, C. J., & Braun, A. R. (2008). Neural substrates of spontaneous musical performance: An fMRI study of jazz improvisation. *PLoS ONE*, *3*, e1679.

Lindenberger, U., Li, S.-C., Gruber, W., & Müller, V. (2009). Brains swinging in concert: Cortical phase synchronization while playing guitar. *BMC Neuroscience*, *10*, 22.

Linklater, F. (1997). Effects of audio- and videotape models on performance achievement of beginning clarinetists. *Journal of Research in Music Education*, *45*, 402–414.

Loehr, J. D., & Palmer, C. (2007). Cognitive and biomechanical influences in pianists' finger tapping. *Experimental Brain Research*, *178*, 518–528.

London, J. (2004). *Hearing in time: Psychological aspects of musical meter*. New York: Oxford.

Longuet-Higgins, H. C., & Lee, C. S. (1982). The perception of musical rhythms. *Perception*, *11*, 115–128.

Lucas, G. (Director), & Williams, J. (Original score). (1977). *Star wars* [Motion picture]. United States: Lucasfilm.

Luck, G., & Sloboda, J. A. (2008). Exploring the spatio-temporal properties of simple conducting gestures using a synchronization task. *Music Perception*, *25*, 225–239.

Luck, G., & Sloboda, J. A. (2009). Spatio-temporal cues for visually mediated synchronization. *Music Perception*, *26*, 465–473.

Luquet, G. (1913). *Les dessins d'un enfant*. [The drawings of a child.] Paris: Alcan.

Lynch, M. P., & Eilers, R. E. (1992). A study of perceptual development for musical tuning. *Perception & Psychophysics*, *52*, 599–608.

Lynch, M. P., Eilers, R. E., Oller, D. K., & Urbano, R. C. (1990). Innateness, experience, and music perception. *Psychological Science*, *1*, 272–276.

Madsen, C. K. (2004). A 30-year follow-up study of actual applied music practice versus estimated practice. *Journal of Research in Music Education*, *52*, 77–88.

Magill, J. M., & Pressing, J. L. (1997). Asymmetric cognitive clock structures in West African rhythm. *Music Perception*, *15*, 189–222.

Maidhof, C., Rieger, M., Prinz, W., & Koelsch, S. (2009). Nobody is perfect: ERP effects prior to performance errors in musicians indicate fast monitoring processes. *PLoS One*, *4*(4).

Malbrán, S. (2000). The development of pulse synchrony: An exploratory study of three-year-old children. *Bulletin of the Council for Research in Music Education*, *147*, 109–115.

Mangione, S., & Nieman, L. Z. (1997). Cardiac auscultatory skills of internal medicine and family practice trainees. A comparison of diagnostic proficiency. *Journal of the American Medical Association*, *278*, 717–722.

Maquet, P., Laureys, S., Perrin, F., Ruby, P., Melchior, G., Boly, M., Vu, T. D., Desseilles, M., & Peigneux, P. (2003). Festina lente: Evidences for fast and slow learning processes and a role for sleep in human motor skill learning. *Learning & Memory*, *10*, 237–239.

Margulis, E. H. (2005). A model of melodic expectation. *Music Perception*, *22*, 663–714.

Margulis, E. H. (2007). Silences in music are musical not silent: An exploratory study of context effects on the experience of musical pauses. *Music Perception*, *24*, 485–506.

Marin, O. S. M., & Perry, D. W. (1999). Neurological aspects of music perception and performance. In D. Deutsch (Ed.), *The psychology of music* (2nd ed., pp. 653–724). San Diego, CA: Academic Press.

Marshall, S. K., & Cohen, A. J. (1988). Effects of musical soundtracks on attitudes toward animated geometric figures. *Music Perception*, *6*, 95–112.

Maruyama, S. (2007). The embodied sense of music: Case studies on the rhetorical function of bodily gestures. *Cognitive Studies*, *14*, 471–493.

Matthews, J. (1999). *The art of childhood and adolescence: The construction of meaning*. London: Routledge.

McAuley, J. D., & Jones, M. R. (2003). Modeling effects of rhythmic context on perceived duration: A comparison of interval and entrainment approaches to short-interval timing. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 1102–1125.

McDermott, J., & Hauser, M. D. (2004). Are consonant intervals music to their ears? Spontaneous acoustic preferences in a nonhuman primate. *Cognition*, *94*, B11–B21.

McDermott, J., & Hauser, M. (2005). The origins of music: Innateness, uniqueness, and evolution. *Music Perception*, *23*, 29–60.

McDermott, J., & Hauser, M. D. (2007). Nonhuman primates prefer slow tempos but dislike music overall. *Cognition*, *104*, 654–668.

MacDonald, R. A. R., Miell, D., & Mitchell, L. (2002). An investigation of children's musical collaborations: The effect of friendship and age. *Psychology of Music*, *30*, 148–163.

McKeage, K. M. (2004). Gender and participation in high school and college instrumental jazz ensembles. *Journal of Research in Music Education*, *52*, 343–356.

McLeod, P., Plunkett, K., & Rolls, E. T. (1998). *Introduction to connectionist modeling of cognitive processes*. Oxford: Oxford University Press.

McMullen, E., & Saffran, J. R. (2004). Music and language: A developmental comparison. *Music Perception*, *21*, 289–311.

McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition*. London: Routledge.

McPherson, G. E. (1994). Factors and abilities influencing sightreading skill in music. *Journal of Research in Music Education*, *42*, 217–231.

McPherson, G. E. (2005). From child to musician: Skill development during the beginning stages of learning an instrument. *Psychology of Music*, *33*, 5–35.

McPherson, G. E. (2009). The role of parents in children's musical development. *Psychology of Music*, *37*, 91–110.

McPherson, G. E., & McCormick, J. (1999). Motivational and self-regulated learning components of musical practice. *Bulletin of the Council for Research in Music Education*, *141*, 98–102.

McPherson, G. E., & Renwick, J. M. (2001). A longitudinal study of self-regulation in children's musical practice. *Music Education Research*, *3*, 169–186.

Meyer, L. B. (1956). *Emotion and meaning in music*. Chicago: Chicago University Press.

Meyer, R., Palmer, C., & Mazo, M. (1998). Affective and coherence responses to Russian laments. *Music Perception*, *16*, 135–150.

Miklaszewski, K. (1989). A case study of a pianist preparing a musical performance. *Psychology of Music*, *17*, 95–109.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97.

Mishra, J. (2002). A qualitative analysis of strategies employed in efficient and inefficient memorization. *Bulletin of the Council for Research in Music Education*, *152*, 74–86.

Mithen, S. (2006). *The singing Neanderthals: The origins of music, language, mind and body*. Cambridge, MA: Harvard University Press.

Mitterschiffthaler, M. T., Fu, C. H. Y., Dalton, J. A., Andrew, C. M., & Williams

S. C. R. (2007). A functional MRI study of happy and sad affective states induced by classical music. *Human Brain Mapping*, *28*, 1150–1162.
Miyazaki, K. (2004). Recognition of transposed melodies by absolute-pitch possessors. *Japanese Psychological Research*, *46*, 270–282.
Moog, H. (1976a). *The musical experience of the pre-school child* (C. Clarke, Trans.). London: Schott.
Moog, H. (1976b). The development of musical experience in children of pre-school age. *Psychology of Music*, *4*, 38–45.
Moore, B. J. C. (2003). *An introduction to the psychology of hearing* (5th ed.). San Diego, CA: Academic Press.
Moorhead, G. E., & Pond, D. (1941). *Music of young children: I. Chant*. Santa Barbara, CA: Pillsbury Foundation for Advancement of Music Education.
Moorhead, G. E., & Pond, D. (1942). *Music of young children: II. General observations*. Santa Barbara, CA: Pillsbury Foundation.
Moorhead, G. E., Sandvik, F., & Wight, D. (1951). *Music of young children: IV. Free use of instruments for musical growth*. Santa Barbara, CA: Pillsbury Foundation.
Morens, D. M. (1995). Mass fainting at medieval rock concerts. *New England Journal of Medicine*, *333*, 1361.
Morrison, S. J., & Fyk, J. (2002). Intonation. In R. Parncutt & G. E. McPherson (Eds.), *The science and psychology of music performance* (pp. 183–197). Oxford: Oxford University Press.
Morrongiello, B. A., Fenwick, K. D., Hillier, L., & Chance, G. (1994). Sound localization in newborn human infants. *Developmental Psychobiology*, *27*, 519–538.
Morrongiello, B. A., & Rocca, P. T. (1987). Infants' localization of sounds in the horizontal plane: Effects of auditory and visual cues. *Child Development*, *58*, 918–927.
Moscovici, S. (1981). *La psychologie de la foule* [The psychology of the crowd]. Paris: Presses Universitaires de France.
Muir, D., & Field, J. (1979). Newborn infants orient to sounds. *Child Development*, *50*, 431–436.
Mürbe, D., Pabst, F., Hofmann, G., & Sundberg, J. (2003). Effects of a professional solo singer education on auditory and kinesthetic feedback – A longitudinal study of singers' pitch control. *Journal of Voice*, *18*, 236–241.
Murnighan, J. K., & Conlon, D. E. (1991). The dynamics of intense work groups: A study of British string quartets. *Administrative Science Quarterly*, *36*, 165–186.
Mutschler, I., Schultze-Bonhage, A., Glauche, V., Demandt, E., Speck, O., & Ball, T. (2007). A rapid sound-action association effect in human insular cortex. *PLoS One*, *2*, e259.
Nager, W., Kohlmetz, C., Altenmüller, E., Rodriguez-Fornells, A., & Münte, T. F. (2003). The fate of sounds in conductors' brains: An ERP study. *Cognitive Brain Research*, *17*, 83–93.
Nakata, T., & Trehub, S. E. (2004). Infants' responsiveness to maternal speech and singing. *Infant Behavior & Development*, *27*, 455–464.
Nantais, K. M., & Schellenberg, E. G. (1999). The Mozart effect: An artifact of preference. *Psychological Science*, *10*, 370–373.
Narmour, E. (1990). *The analysis and cognition of basic melodic structures: The Implication-Realization model*. Chicago: University of Chicago Press.
Narmour, E. (1991). The top-down and bottom-up systems of musical implication: Building on Meyer's theory of emotional syntax. *Music Perception*, *9*, 1–26.

Narmour, E. (1992). *The analysis and cognition of melodic complexity: The Implication-Realization model*. Chicago: University of Chicago Press.

Nettl, B. (2001). Studying music of the world's cultures. In B. Nettl, C. Capwell, I. K. F. Wong, & T. Turino (Eds.), *Excursions in world music* (pp. 1–18). Upper Saddle River, NJ: Prentice-Hall.

Nöcker-Ribaupierre, M. (2004). *Music therapy for premature and newborn infants*. Gilsum, NH: Barcelona Publishers.

North, A. C., Colley, A. M., & Hargreaves, D. J. (2003). Adolescents' perceptions of the music of male and female composers. *Psychology of Music*, *31*, 139–154.

North, A. C., & Hargreaves, D. J. (1998). The effect of music on atmosphere and purchase intentions in a cafeteria. *Journal of Applied Social Psychology*, *28*, 2254–2273.

North, A. C., & Hargreaves, D. J. (2005). Musical communication in commercial contexts. In D. Miell, R. MacDonald, & D. J. Hargreaves (Eds.), *Musical communication* (pp. 405–422). Oxford: Oxford University Press.

North, A. C., & Hargreaves, D. J. (Eds.) (2008). *The social and applied psychology of music*. Oxford: Oxford University Press.

North, A. C., Hargreaves, D. J., & McKendrick, J. (1997). In-store music affects product choice. *Nature*, *390*, 132.

Olds, C. (1985). Fetal response to music. *Midwives Chronicle and Nursing*, *98*, 202–203.

Olsho, L. W. (1984). Infant frequency discrimination. *Infant Behavior & Development*, *7*, 27–35.

O'Gorman, S. (2007). Infant-directed singing in neonatal and paediatric intensive care. *Australian and New Zealand Journal of Family Therapy*, *28*, 100–108.

O'Neill, C. T., Trainor, L. J., & Trehub, S. E. (2001). Infants' responsiveness to fathers' singing. *Music Perception*, *18*, 409–425.

O'Neill, S. A. (1997). Gender and music. In D. J. Hargreaves & A. C. North (Eds.), *The social psychology of music* (pp. 46–66). Oxford: Oxford University Press.

O'Neill, S. A., & Boulton, M. J. (1996). Boys' and girls' preferences for musical instruments. *Psychology of Music*, *24*, 171–183.

Osborne, M. S., & Kenny, D. T. (2008). The role of sensitizing experiences in music performance anxiety in adolescent musicians. *Psychology of Music*, *36*, 447–462.

Osman, A., Albert, R., Ridderinkhof, K. R., Band, G., & van der Molen, M. (2006). The beat goes on: Rhythmic modulation of cortical potentials by imagined tapping. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 986–1005.

Ostri, B., Eller, N., Dahlin, E., & Skylv, G. (1989). Hearing impairment in orchestral musicians. *Scandinavian Audiology*, *18*, 243–249.

Overy, K., Norton, A., Cronin, K., Winner, E., & Schlaug, G. (2005). Examining rhythm and melody processing in young children using fMRI. In G. Avanzini, S. Koelsch, L. Lopez, & M. Majno (Eds.), *The neurosciences and music II: From perception to performance* (pp. 210–218). New York: The New York Academy of Sciences.

Oxendine, J. B. (1984). *Psychology of motor learning* (2nd ed.). New York: Appleton-Century-Crofts.

Ozon, E. (Writer/Director), Delbosc, O., & Missonier, M. (Producers) (2003). *Swimming pool* [Motion picture]. France: Fidélité Productions. (Available from Focus Features, 100 Universal City Plaza, University City, CA 91608)

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 31–346.

Palmer, C. (1997). Music performance. *Annual Review of Psychology*, *48*, 115–138.

Palmer, C., & Krumhansl, C. L. (1987). Independent temporal and pitch structures in determination of musical phrases. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 116–126.

Palmer, C., & Krumhansl, C. L. (1990). Mental representations for musical meter. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 728–741.

Palmer, C., & Pfordresher, P. Q. (2003). Incremental planning in sequence production. *Psychological Review*, *110*, 683–712.

Palmer, C., & van de Sande, C. (1993). Units of knowledge in music performance. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *19*, 457–470.

Palmer, C., & van de Sande, C. (1995). Range of planning in music performance. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 947–962.

Panneton, R. K. (1985). *Prenatal auditory experience with melodies: Effects on postnatal auditory preferences in human newborns*. Unpublished doctoral dissertation, University of North Carolina at Greensboro, North California.

Pantev, C., Engelien, A., Candia, V., & Elbert, T. (2003). Representational cortex in musicians. In I. Peretz & R. Zatorre (Eds.), *The cognitive neuroscience of music* (pp. 382–395). New York: Oxford University Press.

Papoušek, M. (1996). Intuitive parenting: A hidden source of musical stimulation in infancy. In I. Deliège and J. Sloboda (Eds.), *Musical beginnings: Origins and development of musical competence* (pp. 88–112). Oxford: Oxford University Press.

Papoušek, M., Papoušek, H., & Symmes, D. (1991). The meanings of melodies in motherese in tone and stress languages. *Infant Behavior & Development*, *14*, 415–440.

Park, M-Y. (2003). Assessment of potential noise-induced hearing loss with commercial 'Karaoke' noise. *International Journal of Industrial Ergonomics*, *31*, 375–385.

Parncutt, R. (1989). *Harmony: A psychoacoustical approach*. Berlin: Springer-Verlag.

Parncutt, R. (1994). A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception*, *11*, 409–464.

Parsons, L. M., Sergent, J., Hodges, D. A., & Fox, P. T. (2005). The brain basis of piano performance. *Neuropsychologia*, *43*, 119–215.

Pascual-Leone, A. (2003). The brain that makes music and is changed by it. In I. Peretz & R. Zatorre (Eds.), *The cognitive neuroscience of music* (pp. 396–409). New York: Oxford University Press.

Patel, A. D. (2008). *Music, language, and the brain*. New York: Oxford.

Patel, A.D., & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition*, *87*, B35–B45.

Patel, A. D., Iverson, J. R., Bregman, M. R., & Schultz, I. (2009). Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Current Biology*, *19*, 1–4.

Patel, A. D., Peretz, I., Tramo, M. & Labreque, R. (1998). Processing prosodic and musical patterns: A neuropsychological investigation. *Brain and Language*, *61*, 123–144.

Patel, A. D., Wong, M., Foxton, J., Lochy, A., & Peretz, I. (2008). Speech intonation perception deficits in musical tone deafness (congenital amusia). *Music Perception*, *25*, 357–368.

Peretz, I. (1990). Processing of local and global musical information by unilateral brain-damaged patients. *Brain*, *113*, 1185–1205.

Peretz, I. (1993). Auditory atonalia for melodies. *Cognitive Neuropsychology*, *10*, 21–56.

Peretz, I. (2001). Listen to the brain: A biological perspective on musical emotions. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 105–134). Oxford: Oxford University Press.

Peretz, I. (2003). Brain specialization for music: new evidence from congenital amusia. In I. Peretz & R. Zatorre (Eds.), *The cognitive neuroscience of music* (pp. 192–203). New York: Oxford University Press.

Peretz, I., Ayotte, J., Zatorre, R., Mehler, J., Ahad, P., Penhune, V. & Jutras, B. (2002) Congenital amusia: A disorder of fine-grained pitch discrimination. *Neuron*, *33*, 185–191.

Peretz, I., Battico, E., & Tervaniemi, M. (2005). Abnormal electrical brain responses to pitch in congenital amusia. *Annals of Neurology*, *58*, 478–482.

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, *6*, 688–691.

Peretz, I., & Gagnon, L. (1999). Dissociation between recognition and emotional judgment for melodies. *Neurocase*, *5*, 21–30.

Peretz, I., Gagnon, L. & Bouchard, B. (1998). Music and emotion: Perceptual determinants, immediacy and isolation after brain damage. *Cognition*, *68*, 111–141.

Peretz, I., & Zatorre, R. (2003). *The cognitive neuroscience of music*. New York: Oxford.

Peretz, I., & Zatorre, R. J. (2005). Brain organization for music processing. *Annual Review of Psychology*, *56*, 89–114.

Perlman, M., & Krumhansl, C. L. (1996). An experimental study of internal interval standards in Javanese and Western Musicians. *Music Perception*, *14*, 95–116.

Perry, D. W. (1993). A cognitive neuropsychological analysis of melody recall. *Contemporary Music Review*, *9*, 97–111.

Perry, D. W., Zatorre, R. J., Petrides, M., Alivisatos, B., Meyer, E., & Evans, A. C. (1999). Localization of cerebral activity during simple singing. *NeuroReport*, *10*, 3979–3984.

Petacchi, A., Laird, A. R., Fox, P. T., & Bower, J. M. (2005). Cerebellum and auditory function: An ALE meta-analysis of functional neuroimaging studies. *Human Brain Mapping*, *25*, 105–117.

Pfordresher, P. Q. (2003). Auditory feedback in music performance: Evidence for a dissociation of sequencing and timing. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 949–964.

Pfordresher, P. Q. (2006). Coordination of perception and action in music performance. *Advances in Cognitive Psychology*, *2*, 183–198.

Pfordresher, P. Q., & Brown, S. (2007). Poor-pitch singing in the absence of 'tone deafness.' *Music Perception*, *25*, 95–115.

Phillips-Silver, J., & Trainor, L. J. (2005). Feeling the beat: Movement influences infant rhythm perception. *Science*, *308*, 1430.

Phillips-Silver, J., & Trainor, L. J. (2007). Hearing what the body feels: Auditory encoding of rhythmic movement. *Cognition*, *105*, 533–546.

Piaget, J. (1952). *The origins of intelligence in children*. New York: International Universities Press. (Original work published 1936)

Pick, A. D., Gross, D., Heinrichs, M., Love, M., & Palmer, C. (1994). Development of perception of the unity of musical events. *Cognitive Development*, *9*, 355–375.

Pickering, S., & Repacholi, B. (2001). Modifying children's gender-typed musical instrument preferences: The effects of gender and age. *Sex Roles*, *45*, 623–643.

Pickles, J. O. (2008) *An introduction to the physiology of hearing* (3rd ed.). Bingley: Emerald Books.

Pierce, J. R. (1992). *The science of musical sound*. New York: Freeman.

Pinker, S. (1997). *How the mind works*. New York: W. W. Norton.

Plack, C. J. (2005). *The sense of hearing*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Plantinga, J., & Trainor, L. J. (2004). Memory for melody: Infants use a relative pitch code. *Cognition*, *98*, 1–11.

Plomp, R., & Levelt, W. J. M. (1965). Tonal consonance and critical bandwidth. *Journal of the Acoustical Society of America*, *38*, 548–560.

Poggi, I. (2002). The lexicon of the conductor's face. In P. McKevitt, S. Ó. Nulláin, & C. Mulvihill (Eds.), *Language, vision, and music: Selected papers from the 8th International Workshop on the Cognitive Science of Natural Language Processing, Galway, Ireland 1999. Advances in consciousness research* (pp. 271–284). Amsterdam: John Benjamins.

Povel, D. J. (1981). Internal representation of simple temporal patterns. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 3–18.

Povel, D. J., & Essens, P. (1985). Perception of temporal patterns. *Music Perception*, *2*, 411–441.

Pratt, C. C. (1931). *The meaning of music: A study in psychological aesthetics* (1st ed.). New York and London: McGraw-Hill Book Company, Inc.

Pressing, J. (1999). The referential dynamics of cognition and action. *Psychological Review*, *106*, 714–747.

Pujol, R., & Lavigne-Rebillard, M. (1985). Early stages of innervation and sensory cell differentiation in the human fetal organ of Corti. *Acta Otolaryngol*, *424*, 43–50.

Pujol, R., Lavigne-Rebillard, M., & Uziel, A. (1990). Physiological correlates of development of the human cochlea. *Seminars in Perinatology*, *14*, 275–280.

Querleu, D., Lefebre, C., Renard, X., Titran, M., Morillion, M., & Crépin, G. (1984). Réactivité du nouveau-né de moins de deux heures de vie a la voix maternelle [Responsivity of the newborn to the maternal voice at less than two hours old]. *Journal de Gynecologie, Obstétrique et Biologie de la Reproduction*, *13*, 125–134.

Querleu, D., Renard, X., Boutteville, C., & Crépin, G. (1989). Hearing by the human fetus? *Seminars in Perinatology*, *13*, 409–420.

Querleu, D., Renard, X., Versyp, F., Paris-Delrue, L., & Crépin, G. (1988). Fetal hearing. *European Journal of Obstetrics and Gynecology and Reproductive Biology*, *29*, 191–212.

Radocy, R. E., & Boyle, J. D. (2003). *Psychological foundations of musical behavior*. Springfield, IL: Thomas.

Rainbow, E. L. (1981). A final report on a three-year investigation of the rhythmic abilities of preschool-aged children. *Bulletin of the Council for Research in Music Education*, *66–67*, 69–73.

Ramnani, N., & Passingham, R. E. (2001). Changes in the human brain during rhythm learning. *Journal of Cognitive Neuroscience*, *13*, 952–966.
Rasch, R., & Plomp, R. (1999). The perception of musical tones. In D. Deutsch (Ed.), *The psychology of music* (pp. 89–112). San Diego, CA: Academic Press.
Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of 'what' and 'where' in auditory cortex. *Proceedings of the National Academy of Sciences*, *97*, 11800–11806.
Rauscher, F. H., Shaw, G. L., & Ky, K. N. (1993). Music and spatial task performance. *Nature*, *365*, 6447.
Rauscher, F. H., Shaw, G. L., & Ky, K. N. (1995). Listening to Mozart enhances spatial-temporal reasoning: Towards a neurophysiological basis. *Neuroscience Letters*, *185*, 44–47.
Reichardt, W., Alim, A., & Schmidt, W. (1974). Abhängigkeit der Grenzen zwischen brauchbarer und unbrauchbarer Durchsichtigkeit von der Art des Musikmotives, der Nachhallzeit und der Nachhalleinsatzzeit. *Applied Acoustics*, *7*, 243–264.
Repp, B. H. (1991). Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's 'Träumerei.' *Journal of the Acoustical Society of America*, *92*, 2546–2568.
Repp, B. H. (1999). Effects of auditory feedback deprivation on expressive piano performance. *Music Perception*, *16*, 409–438.
Repp, B. H. (2000). Pattern typicality and dimensional interactions in pianists' imitation of expressive timing and dynamics. *Music Perception*, *18*, 173–212.
Révész, G. (1954). *Introduction to the psychology of music*. (C. I. C. de Courcy, Trans.). Norman, OK: University of Oklahoma.
Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169–192.
Roballey, T. C., McGreevy, C., Rongo, R. R., Schwantes, M. L., Steger, P. J., Wininger, M. A., & Gardner, E. B. (1985). The effect of music on eating behavior. *Bulletin of the Psychonomic Society*, *23*, 221–222.
Robbe-Grillet, A. (Director), Halfon, S., & Tomaskovic, J. (Producers) (1968). *L'Homme qui ment* [The man who lies] [Motion picture]. France.
Robertson, S. D., Zelaznik, H. N., Lantero, D., Bojczyk, K. G., Spencer, R. M., Doffin, J. G., & Schneidt, T. (1999). Correlations for timing consistency among tapping and drawing tasks: Evidence against a single timing process for motor control. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1316–1330.
Rock, A. M. L., Trainor, L. J., & Addison, T. L. (1999). Distinctive messages in infant-directed lullabies and play songs. *Developmental Psychology*, *35*, 527–534.
Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. Oxford: Oxford University Press.
Rosenthal, R. K. (1984). Relative effects of guided model, model only, guide only, and practice only treatments on the accuracy of advanced instrumentalists' musical performance. *Journal of Research in Music Education*, *3*, 265–273.
Rosenthal, R. K., Wilson, M., Evans, M., & Greenwalt, L. (1988). Effects of different practice conditions on advanced instrumentalists' performance accuracy. *Journal of Research in Music Education*, *36*, 250–257.
Rosner, B. S., & Meyer, L. B. (1982). Melodic processes and the perception of music. In D. Deutsch (Ed.), *The psychology of music* (pp. 317–343). San Diego, CA: Academic Press.

Ross, S. L. (1985). The effectiveness of mental practice in improving the performance of college trombonists. *Journal of Research in Music Education*, *33*, 221–230.

Rubin-Rabson, G. (1940a). Studies in the psychology of memorizing music: II. A comparison of massed and distributed learning. *Journal of Educational Psychology*, *31*, 270–284.

Rubin-Rabson, G. (1940b). Studies in the psychology of memorizing music: III. A comparison of the whole and the part approach. *Journal of Educational Psychology*, *31*, 460–476.

Rubin-Rabson, G. (1941a). Studies in the psychology of memorizing music: V. A comparison of pre-study periods of varied length. *Journal of Educational Psychology*, *32*, 101–112.

Rubin-Rabson, G. (1941b). Studies in the psychology of memorizing music: VI. A comparison of two forms of mental rehearsal and keyboard overlearning. *Journal of Educational Psychology*, *32*, 593–602.

Ruiz, M. H., Jabusch, H. C., & Altenmüller, E. (2009). Detecting wrong notes in advance: Neuronal correlates of error monitoring in pianists. *Cerebral Cortex*, *19*(11), 2625–2639.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*, 1161–1178.

Rutkowski, J. (1997). The nature of children's singing voices. In B. A. Roberts (Ed.), *The phenomenon of singing* (pp. 201–209). St. John's, NF: Memorial University Press.

Sabine, W. C. (1922). *Collected papers on acoustics*. Cambridge, MA: Harvard University Press.

Sacks, O. (2007). *Musicophilia: Tales of music and the brain*. New York: Knopf.

Saffran, J. R., & Griepentrog, G. J. (2001). Absolute pitch in infant auditory learning: Evidence for developmental reorganization. *Developmental Psychology*, *37*, 74–85.

Saffran, J. R., Loman, M. M., & Robertson, R. R. W. (2000). Infant memory for musical experiences. *Cognition*, *77*, B15–B23.

Sakai, K., Hikosaka, O., Miyauchi, S., Takino, R., Tamada, T., Iwata, N. K., & Nielsen, M. (1999). Neural representation of rhythm depends on its interval ratio. *Journal of Neuroscience*, *19*, 10074–10081.

Saldanha, F. I., & Corso, J. F. (1964). Timbre cues and the identification of musical instruments. *Journal of the Acoustical Society of America*, *36*, 2021–2026.

Salk, L. (1961). Mothers' heartbeat as an imprinting stimulus. *Transactions of the New York Academy of Sciences, Series 2*, *4*, 753–763.

Samson, S., & Ehrlé, N. (2003). Cerebral substrates for musical temporal processes. In I. Peretz & R. Zatorre (Eds.), *The cognitive neuroscience of music* (pp. 204–216). New York: Oxford University Press.

Santesso, D. L., Schmidt, L. A., & Trainor, L. J. (2007). Frontal brain electrical activity (EEG) and heart rate in response to affective infant-directed (ID) speech in 9-month-old infants. *Brain and Cognition*, *65*, 14–21.

Sataloff, R. T. (1991). Hearing loss in musicians. *American Journal of Otology*, *12*, 122–127.

Satoh, M., Takeda, K., & Kuzuhara, S. (2007). A case of auditory agnosia with impairment of perception and expression of music: Cognitive processing of tonality. *European Neurology*, *58*, 70–77.

Satt, B. J. (1984). *An investigation into the acoustical induction of intra-uterine learning*.

Unpublished doctoral dissertation. Los Angeles: California School of Professional Psychologists.
Schellenberg, E. G. (1996). Expectancy in melody: Tests of the implication-realization model. *Cognition*, *58*, 75–125.
Schellenberg, E. G. (1997). Simplifying the implication-realization model of melodic expectancy. *Music Perception*, *14*, 295–318.
Schellenberg, E. G. (2006). Long-term positive associations between music lessons and IQ. *Journal of Educational Psychology*, *98*, 457–468.
Schellenberg, E. G., Iverson, P., & McKinnon, M. C. (1999). Name that tune: Identifying popular recordings from brief excerpts. *Psychonomic Bulletin & Review*, *64*, 641–646.
Schellenberg, E. G., & Trehub, S. E. (1996). Natural musical intervals: Evidence from infant listeners. *Psychological Science*, *7*, 272–277.
Schellenberg, E. G., & Trehub, S. E. (2003). Good pitch memory is widespread. *Psychological Science*, *14*, 262–266.
Schenker, H. (1935). *Free Composition (Der Freie Satz)*. (E. Oster, Trans.). New York: Longman.
Schenker, H. (1954). *Harmony*. (E. M. Borgese, Trans.). Chicago: Chicago University Press. (Original work published 1906)
Schlaug, G. (2009). Music, musicians, and brain plasticity. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (pp. 197–207). New York: Oxford University Press.
Schlaug, G., Jäncke, L., Huang, Y., & Steinmetz, H. (1995, February 3). In vivo evidence of structural brain asymmetry in musicians. *Science*, *267*, 699–701.
Schlaug, G., Jäncke, L., Huang, Y., Staiger, J. F., & Steinmetz, H. (1995). Increased corpus callosum size in musicians. *Neuropsychologia*, *33*, 1047–1055.
Schlaug, G., Norton, A., Overy, K., & Winner, E. (2005). Effects of music training on the child's brain and cognitive development. In G. Avanzini, S. Koelsch, L. Lopez, & M. Majno (Eds.), *The neurosciences and music II: From perception to performance* (pp. 219–230). New York: The New York Academy of Sciences.
Schleuter, S. L. (1997). *A sound approach to teaching instrumentalists* (2nd ed.). Kent, OH: Kent State University Press.
Schouten, J. F. (1938). The perception of subjective tones. *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen* [Proceedings of the Royal Dutch Academy of Knowledge], *34*, 1418–1424.
Schubert, E., & Stevens, C. (2006). The effect of implied harmony, contour and musical expertise on judgments of similarity of familiar melodies. *Journal of New Music Research*, *35*, 161–174.
Schutz, M., & Kubovy, M. (2009). Deconstructing a musical illusion: Point-light representations capture salient properties of impact motions. *Canadian Acoustics*, *37*, 23–28.
Schutz, M., & Lipscomb, S. (2007). Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception*, *36*, 888–897.
Scruton, R. (1983). *The aesthetic understanding*. Manchester: Carcanet.
Seashore, C. E. (1919). *The psychology of musical talent*. Boston: Silver Burdett.
Seashore, C. E. (1938). *The psychology of music*. New York: McGraw-Hill.
Seppänen, M., Brattico, E., & Tervaniemi, M. (2007). Practice strategies of musicians modulate neural processing and the learning of sound patterns. *Neurobiology of Learning and Memory*, *87*, 236–247.

Sergent, J., Zuck, E., Terriah, S., & MacDonald, B. (1992, July 3). Distributed neural network underlying musical sight-reading and keyboard performance. *Science*, *257*, 106–109.

Seuss, T.G. (1957). *The cat in the hat*. New York: Beginner Books/Random House.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial-ordering mechanism in sentence production. In W. E. Cooper & C. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 295–342). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Shelley, S. (1981). Investigating the musical capabilities of young children. *Bulletin of the Council for Research in Music Education*, *68*, 26–34.

Shenfield, T., Trehub, S. E., & Nakata, T. (2003). Maternal singing modulates infant arousal. *Psychology of Music*, *31*, 365–375.

Shepard, R. N. (1982). Structural representations of musical pitch. In D. Deutsch (Ed.), *The psychology of music* (pp. 344–390). New York: Academic Press.

Shetler, D. J. (1989). The inquiry into prenatal musical experience. *Pre- and Peri-Natal Psychology*, *3*, 171–189.

Shuter-Dyson, R. & Gabriel, C. (1981). *The psychology of musical ability*. London: Methuen.

Simon, H. A., & Chase, W. G. (1973). Skill in chess. *American Scientist*, *61*, 394–403.

Simonton, D. K. (1991). Emergence and realization of genius: The lives and works of 120 classical composers. *Journal of Personality and Social Psychology*, *61*, 829–840.

Sloboda, J. [A.] (1974). The eye-hand span: An approach to the study of sight reading. *Psychology of Music*, *2*, 4–10.

Sloboda, J. A. (1985). *The musical mind: The cognitive psychology of music*. Oxford: Oxford University Press.

Sloboda, J. A. (1991). Music structure and emotional response: Some empirical findings. *Psychology of Music*, *19*, 110–120.

Sloboda, J. A. (2005). *Exploring the musical mind: Cognition, emotion, ability, function*. Oxford: Oxford University Press.

Sloboda, J. A., Davidson, J. W., Howe, M. J. A., & Moore, D. G. (1996). The role of practice in the development of performing musicians. *British Journal of Psychology*, *87*, 287–309.

Sloboda, J. A., & Gregory, A. H. (1980). The psychological reality of musical segments. *Canadian Journal of Experimental Psychology*, *34*, 274–280.

Sloboda, J. A., & O'Neill, S. A. (2001). Emotions in everyday listening to music. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 415–430). Oxford: Oxford University Press.

Sloboda, J. A., Wise, K. J., & Peretz, I. (2005). Quantifying tone deafness in the general population. *Annals of the New York Academy of Sciences*, *1060*, 255–261.

Smith, G. M. (1999). Movie music as moving music: Emotion, cognition, and the film score. In C. Plantinga & G. M. Smith (Eds.), *Passionate views: Film, cognition, and emotion* (pp. 146–167). Baltimore: Johns Hopkins University Press.

Smith, J. D. (1997). The place of novices in music science. *Music Perception*, *14*, 227–262.

Smithrim, K., & Upitis, R (2007). *Listen to their voices: Research and practice in early childhood music*. Toronto: Canadian Music Educators' Association.

Snow, C. W., & McGaha, C. G. (2003). *Infant development* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.

Snyder, B. (2000). *Music and memory: An introduction*. Cambridge, MA: MIT Press.
Snyder, J. S., & Large, E. W. (2005). Gamma-band activity reflects the metric structure of rhythmic tone sequences. *Cognitive Brain Research*, *24*, 117–126.
Sontag, L. W., & Wallace, R. F. (1935). The movement response of the human fetus to sound stimuli. *Child Development*, *6*, 253–258.
Sosniak, L. A. (1985). Learning to be a concert pianist. In B. S. Bloom (Ed.), *Developing talent in young people* (pp. 19–67). New York: Ballantine.
Spelke, E. S. (1976). Infants' intermodal perception of events. *Cognitive Psychology*, *8*, 553–560.
Sperry, R. W., Gazzaniga, M. S., & Bogen, J. E. (1969). Interhemispheric relationships: The neocortical commissures; syndromes of hemisphere disconnection. In P. J. Vinken & G. W. Bruyn (Eds.), *Handbook of clinical neurology* (Vol. 4, pp. 273–290). Amsterdam: North-Holland Publishing Company, and New York: John Wiley and Sons.
Spielberg, S. (Director), & Williams, J. (Original score). (1998). *Saving Private Ryan* [Motion picture]. United States: Amblin Entertainment.
Spielberg, S. (Director), Goldman, G., Shusett, R. (Executive Producers), & Williams, J. (Original score). (2002). *Minority report* [Motion picture]. United States: DreamWorks LLC and Twentieth Century Fox.
Spinoza, B. (1677) *Ethics: Part I, Propositions IV; Part II, Propositions XXVII, XXVIII*. (A. Boyle and G. H. R. Parkinson, Trans., 1987). London: Dent.
Stack, D., Muir, D., Sherriff, F., & Roman, J. (1989). Development of infant reaching in the dark to luminous objects and 'invisible sounds.' *Perception*, *18*, 69–82.
Standley, J. M. (1999). Music therapy in the NICU: Pacifier-activated-lullabies (PAL) for reinforcement of non-nutritive sucking. *International Journal of Arts Medicine*, *6*, 17–21.
Standley, J. M. (2003). *Music therapy with premature infants: Research and developmental interventions*. Silver Spring, MD: The American Music Therapy Association.
Starr, A., Amlie, R. N., Martin, W. H., & Sanders, S. (1977). Development of auditory function in newborn infants revealed by auditory brainstem potentials. *Pediatrics*, *60*, 831–839.
Steblin, R. (1995). The gender stereotyping of musical instruments in the Western tradition. *Canadian University Music Review*, *16*, 128–144.
Steptoe, A. (2001). Negative emotions in music making: The problem of performance anxiety. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 291–307). New York: Oxford University Press.
Stevens, C., & Byron, T. (2009). Universals in music processing. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (pp. 14–23). Oxford: Oxford University Press.
Stevens, S. S. (1936). A scale for the measurement of psychological magnitude: Loudness. *Psychological Review*, *43*, 405–416.
Sundberg, J. (1982). Perception of singing. In D. Deutsch (Ed.), *The psychology of music*. New York: Academic.
Sundberg, J. (1987). *The science of the singing voice*. Dekalb, IL: Northern Illinois University Press.
Swanwick, K. (1994). *Musical knowledge: Intuition, analysis, and music education*. London: Routledge.

Swanwick, K., & Tillman, J. (1986). The sequence of musical development: A study of children's composition. *British Journal of Music Education*, *3*, 305–339.

Takeuchi, A. H., & Hulse, S. H. (1993). Absolute pitch. *Psychological Bulletin*, *113*, 345–361.

Tan, S. L., & Spackman, M. P. (2005). Listeners' judgments of the musical unity of structurally altered and intact musical compositions. *Psychology of Music*, *33*, 133–153.

Tan, S. L., Spackman, M. P., & Bezdek, M. A. (2007). Viewers' interpretations of film characters' emotions: Effects of presenting film music before or after a character is shown. *Music Perception*, *25*, 135–152.

Tan, S. L., Spackman, M. P., & Peaslee, C. L. (2006). Effects of repeated exposure on liking and judgments of thematic unity of patchwork and intact compositions. *Music Perception*, *23*, 407–421.

Tan, S. L., Spackman, M. P., & Wakefield, E. M. (2008). Effects of diegetic and non-diegetic presentation of film music on viewers' interpretation of film narrative. *Conference Proceedings for the 2008 International Conference of Music Perception and Cognition* (pp. 588–593), Hokkaido University, Japan. Australia: Causal Productions. [Journal publication forthcoming]

Tarnowski, S. M. (1999). Musical play and young children. *Music Educators Journal*, *86*, 26–29.

Temperley, D. (2000). Meter and grouping in African music: A view from music theory. *Ethnomusicology*, *44*, 65–96.

Temperley, D. (2001). *The cognition of basic musical structures*. Cambridge, MA: MIT Press.

Tenzer, M. (1991). *Balinese music*. Singapore: Periplus.

Tenzer, M. (2000). *Gamelan gong kebyar*. Chicago: University of Chicago Press.

Terhardt, E. (1974). Pitch, consonance, and harmony. *Journal of the Acoustical Society of America*, *55*, 1061–1069.

Terwogt, M. M., & van Grinsven, F. (1991). Musical expression of moodstates. *Psychology of Music*, *19*, 99–109.

Thayer, J. F., & Levenson, R. W. (1983). Effects of music on psychophysiological responses to a stressful film. *Psychomusicology*, *3*, 44–52.

Theiler, A. M., & Lippman, L. G. (1995). Effects of mental practice and modeling on guitar and vocal performance. *Journal of General Psychology*, *122*, 329–343.

Thompson, W. F., Cuddy, L., & Plaus, C. (1997). Expectancies generated by melodic intervals: Evaluation of principles of melodic implication in a melody-completion task. *Perception & Psychophysics*, *59*, 1069–1076.

Thompson, W. F., Russo, F. A., & Sinclair, D. (1994). Effects of underscoring on the perception of closure in filmed events. *Psychomusicology*, *13*, 9–27.

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2001). Arousal, mood, and the Mozart effect. *Psychological Science*, *12*, 248–251.

Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of tonality: A self-organizing approach. *Psychological Review*, *107*, 885–913.

Tillmann, B., & Bigand, E. (2002). A comparative review of priming effects in language and music. In P. McKevitt, S. Ó. Nulláin, & C. Mulvihill (Eds.), *Language, vision, and music: Selected papers from the 8th International Workshop on the Cognitive Science of Natural Language Processing, Galway, Ireland 1999. Advances in consciousness research* (pp. 231–240). Amsterdam: John Benjamins.

Tillmann, B., Janata, P., & Bharucha, J. J. (2003). Activation of the inferior frontal cortex in musical priming. *Cognitive Brain Research*, *16*, 145–161.

Timmers, R., & Ashley, R. (2007). Emotional ornamentation in performances of a Handel sonata. *Music Perception, 25*, 117–134.

Todd, N. [P. M.]. (1985). A model of expressive timing in music performance in tonal music. *Music Perception*, *3*, 33–58.

Tomatis, A. (1996). *The ear and language*. Ontario, Canada: Moulin.

Trainor, L. J. (1996). Infant preferences for infant-directed versus noninfant-directed playsongs and lullabies. *Infant Behavior & Development*, *19*, 83–92.

Trainor, L. J., Tsang, C. D., & Cheung, V. H. W. (2002). Preference for sensory consonance in 2- and 4-month-old infants. *Music Perception*, *20*, 187–194.

Trainor, L. J., Wu, L., & Tsang, C. D. (2004). Long-term memory for music: Infants remember tempo and timbre. *Developmental Science*, *7*, 289–296.

Trehub, S. E. (1993). Temporal auditory processing in infancy. *Annals of the New York Academy of Sciences*, *682*, 137–149.

Trehub, S. (2000). Human processing predispositions and musical universals. In N. L. Wallin, B. Merker, & S. Brown (Eds), *The origins of music* (pp. 427–448). Cambridge, MA: MIT Press.

Trehub, S. E., Bull, D., & Thorpe, L. A. (1984). Infants' perception of melodies: The role of melodic contour. *Child Development*, *55*, 821–830.

Trehub, S. E., Hill, D. S., & Kamenetsky, S. B. (1997). Parents' sung performances for infants. *Canadian Journal of Experimental Psychology*, *51*, 385–396.

Trehub, S. E., & Schellenberg, E. G. (1995). Music: Its relevance to infants. *Annals of Child Development*, *11*, 1–24.

Trehub, S. E., Schellenberg, E. G., & Kamenetsky, S. B. (1999). Infants' and adults' perception of scale structure. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 965–975.

Trehub, S. E., & Thorpe, L. A. (1989). Infants' perception of rhythm: Categorization of auditory sequences by temporal structure. *Canadian Journal of Psychology*, *43*, 217–229.

Trehub, S. E., Thorpe, L. A., & Morrongiello, B. A. (1985). Infants' perception of melodies: Changes in a single tone. *Infant Behavior & Development*, *8*, 213–223.

Trehub, S. E., Unyk, A. M., & Henderson, J. L. (1994). Children's songs to infant siblings: Parallels with speech. *Journal of Child Language*, *21*, 735–744.

Triplett, N. (1898). The dynamogenic factors in pacemaking and competition. *American Journal of Psychology*, *9*, 507–533.

Urmson, J. O. (1973). Representation in music. In G. Vesey (Ed.), *Philosophy and the arts, Royal Institute of Philosophy lectures (Vol. 6)*. London: Macmillan.

Veneklasen, P. S. (1975). Design considerations from the viewpoint of the professional consultant. In R. Mackenzie (Ed.), *Auditorium acoustics: The proceedings of an international symposium on architectural acoustics*. New York: Wiley-Halsted.

Verbinski, G. (Director), & Zimmer, H. (Original score). (2006). *Pirates of the Caribbean: Dead man's chest* [Motion picture]. United States: Walt Disney Pictures.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition*, *101*, 80–113.

Vitouch, O. (2001). When your ear sets the stage: Musical context effects in film perception. *Psychology of Music*, *29*, 70–83.

von Békésy, G. (1960). *Experiments in hearing*. New York: McGraw Hill.

Von Hippel, P. & Huron, D. (2000). Why do skips precede reversals? The effects of tessitura and melodic structure. *Music Perception*, *18*, 59–86.

Vygotsky, L. S. (1978). *Mind in society: The development of higher mental processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Cambridge, MA: Harvard University Press. (Original work published 1930)

Wachowski, A., & Wachowski, L. (Writers/Directors), & Davis, D. (Original score). (2003). *The matrix reloaded* [Motion picture]. USA/Australia: Warner Bros Pictures.

Wachowski, A., & Wachowski, L. (Writers/Directors), & Davis, D. (Original score). (2003). *The matrix revolutions* [Motion picture]. Australia/USA: Warner Bros Pictures.

Wallin, N. L., Merker, B., & Brown, S. (2000). *The origins of music*. Cambridge, MA: MIT Press.

Warren, J. D., Uppenkamp, S., Patterson, R. D., & Griffiths, T. D. (2003). Separating pitch chroma and pitch height in the human brain. *Proceedings of the National Academy of Sciences*, *100*, 10038–10042.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, *167*, 392–393.

Wehr-Flowers, E. (2006). Differences between male and female students' confidence, anexiety, and attitude toward learning jazz improvisation. *Journal of Research in Music Education*, *54*, 337–349.

Welch, G., F. (1985). A schema theory of how children learn to sing in tune. *Psychology of Music*, *13*, 3–18.

Welch, G. F., Howard, D. M., & Rush. C. (1989). Real-time visual feedback in the development of vocal pitch accuracy in singing. *Psychology of Music*, *7*, 146–157.

Werker, J. F., Pegg, J. E., & McLeod, P. J. (1994). A cross-language investigation of infant preference for infant-directed communication. *Infant Behavior & Development*, *17*, 323–333.

Wertheimer, M. (1938). Gestalt theory. In W. D. Ellis (Ed. & Trans.), *A source book of Gestalt psychology* (pp. 1–11). London: Routledge & Kegan Paul. (Original work published 1925)

Wilder, B. (Director), & Rózsa, M. (Original score). (1945). *The lost weekend* [Motion picture]. United States: Paramount.

Wilkin, P. (1995). A comparison of fetal and newborn responses to music and sound stimuli with and without daily exposure to a specific piece of music. *Bulletin of the Council for Research in Music Education*, *127*, 163–169.

Williamon, A., & Valentine, E. (2000). Quantity and quality of musical practice as predictors of performance quality. *British Journal of Psychology*, *91*, 353–376.

Williamon, A., & Valentine, E. (2002). The role of retrieval structures in memorizing music. *Cognitive Psychology*, *44*, 1–32.

Wilson, G. D. (1997). Performance anxiety. In D. J. Hargreaves & A. C. North (Eds.), *The social psychology of music* (pp. 229–245). Oxford: Oxford University Press.

Wilson, G. D., & Roland, D. (2002). Performance anxiety. In R. Parncutt & G. E. McPherson (Eds.), *The science and psychology of music performance* (pp. 47–61). New York: Oxford University Press.

Wing, A. M., & Kristofferson, A. B. (1973). Response delays and the timing of discrete motor responses. *Perception & Psychophysics*, *14*, 5–12.

Winner, E. (2000). The origins and ends of giftedness. *American Psychologist*, *55*, 159–169.

Wise, R. (Director). (1965). *The sound of music* [Motion picture]. United States: Robert Wise Productions.

Wollner, C., & Auhagen, W. (2008). Perceiving conductors' expressive gestures from different visual perspectives. *Music Perception*, *26*, 129–143.

Wong, P. C. M., Parsons, L. M., Martinez, M., & Diehl, R. L. (2004). The role of the insular cortex in pitch pattern perception: The effect of linguistic contexts. *Journal of Neuroscience*, *24*, 9153–9160.

Wood, D., Bruner, J., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, *17*, 89–100.

Woodward, S. C. (1992). Intrauterine rhythm and blues? *British Journal of Obstetrics and Gynaecology*, *99*, 787–790.

Wright, A. A., Rivera, J. J., Hulse, S.H., Shyan, M., & Neiworth, J. J. (2000). Music perception and octave generalization in rhesus monkeys. *Journal of Experimental Psychology: General*, *129*, 291–307.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*, 459–482.

Yost, W. A. (2000). *Fundamentals of hearing* (4th ed.). San Diego, CA: Academic Press.

Young, S. (2003a). Time-space structuring in spontaneous play on educational percussion instruments among three- and four-year-olds. *British Journal of Music Education*, *20*, 45–59.

Young, S. (2003b). The interpersonal dimension: A potential source of musical creativity for young children? *Musicae Scientiae, Special issue: Musical creativity special 10th anniversary conference issue: Award papers*, *2003–2004*, 175–191.

Young, S. (2003c). *Music with the under-fours*. London: Routledge-Falmer.

Zajonc, R. B. (1965). Social facilitation. *Science*, *149*, 261–274.

Zattore, R. (2003). Neural specializations for tonal processing. In I. Peretz & R. Zatorre (Eds.), *The cognitive neuroscience of music* (pp. 231–246). New York: Oxford University Press.

Zatorre, R., & Halpern, A. (2005). Mental concerts: Musical imagery and the auditory cortex. *Neuron*, *47*, 9–12.

Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992, May 8). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, *256*, 846–849.

Zimmer, E. Z., Divon, M. Y., Vilensky, A., Sarna, Z., Peretz, B. A., & Paldi, E. (1982). Maternal exposure to music and fetal activity. *European Journal of Obstetrics and Gynecology and Reproductive Biology*, *13*, 209–213.

Name index

Subject index