SPRINGER HANDBOOK OF AUDITORY RESEARCH

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

SPRINGER HANDBOOK OF AUDITORY RESEARCH

Volume 1: The Mammalian Auditory Pathway: Neuroanatomy
Edited by Douglas B. Webster, Arthur N. Popper, and Richard R. Fay

Volume 2: The Mammalian Auditory Pathway: Neurophysiology
Edited by Arthur N. Popper and Richard R. Fay

Volume 3: Human Psychophysics
Edited by William Yost, Arthur N. Popper, and Richard R. Fay

Volume 4: Comparative Hearing: Mammals
Edited by Richard R. Fay and Arthur N. Popper

Volume 5: Hearing by Bats
Edited by Arthur N. Popper and Richard R. Fay

Volume 6: Auditory Computation
Edited by Harold L. Hawkins, Teresa A. McMullen, Arthur N. Popper, and Richard R. Fay

Volume 7: Clinical Aspects of Hearing
Edited by Thomas R. Van De Water, Arthur N. Popper, and Richard R. Fay

Volume 8: The Cochlea
Edited by Peter Dallos, Arthur N. Popper, and Richard R. Fay

Volume 9: Development of the Auditory System
Edited by Edwin W Rubel, Arthur N. Popper, and Richard R. Fay

Volume 10: Comparative Hearing: Insects
Edited by Ronald Hoy, Arthur N. Popper, and Richard R. Fay

Volume 11: Comparative Hearing: Fish and Amphibians
Edited by Richard R. Fay and Arthur N. Popper

Volume 12: Hearing by Whales and Dolphins
Edited by Whitlow W.L. Au, Arthur N. Popper, and Richard R. Fay

Volume 13: Comparative Hearing: Birds and Reptiles
Edited by Robert Dooling, Arthur N. Popper, and Richard R. Fay

Volume 14: Genetics and Auditory Disorders
Edited by Bronya J.B. Keats, Arthur N. Popper, and Richard R. Fay

Volume 15: Integrative Functions in the Mammalian Auditory Pathway
Edited by Donata Oertel, Richard R. Fay, and Arthur N. Popper

Volume 16: Acoustic Communication
Edited by Andrea Simmons, Arthur N. Popper, and Richard R. Fay

Volume 17: Compression: From Cochlea to Cochlear Implants
Edited by Sid P. Bacon, Richard R. Fay, and Arthur N. Popper

Volume 18: Speech Processing in the Auditory System
Edited by Steven Greenberg, William Ainsworth, Arthur N. Popper, and Richard R. Fay

Volume 19: The Vestibular System
Edited by Stephen M. Highstein, Richard R. Fay, and Arthur N. Popper

Volume 20: Cochlear Implants: Auditory Prostheses and Electric Hearing
Edited by Fan-Gang Zeng, Arthur N. Popper, and Richard R. Fay

Volume 21: Electroreception
Edited by Theodore H. Bullock, Carl D. Hopkins, Arthur N. Popper, and Richard R. Fay

Continued after index

William A. Yost
Arthur N. Popper
Richard R. Fay
Editors

Auditory Perception of Sound Sources

Springer

William A. Yost
Speech and Hearing Sciences
Arizona State University
Tempe, AZ 85287
USA
William.yost@asu.edu

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

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

Series Editors:
Richard R. Fay
Parmly Hearing Institute and Department of Psychology
Loyola University Chicago
Chicago, IL 60626
USA

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

ISBN-13: 978-0-387-71304-5　　e-ISBN-13: 978-0-387-71305-2

Library of Congress Control Number: 2007928313

Printed on acid-free paper.

9 8 7 6 5 4 3 2 1

springer.com

Contents

Series Preface

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

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

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

RICHARD R. FAY, Chicago, IL
ARTHUR N. POPPER, College Park, MD

Volume Preface

To survive, animals must navigate, find food, avoid predators, and reproduce; and many species survive based on their ability to communicate. All of these crucial behaviors allow animals to function in a crowded world of obstacles, objects, and other animals. Many of these objects vibrate and produce sound, and sound may be used to determine the sources of the sound and to serve as a basis for communication. Sounds produced by different sources are combined in one sound field that must be parsed into information that allows for the determination of the individual sources. This process begins at the level of the auditory receptor organ, but is primarily accomplished by processing of the peripheral code in the brain. Given the variety of sources that produce sound, the complexity of the world in which these sources exist, and the lack of peripheral receptors to analyze sound sources per se, determining the sources of sound presents a significant challenge for the auditory system. At present, not a great deal is known about how the auditory system deals with this challenge. This book reviews several topics that are likely relevant to enhance an understanding of the auditory system's ability to determine sound sources.

Yost, in Chapter 1, provides an overview of the volume and the issues that arise in considering sound source perception. Chapter 2, by Lufti, describes the properties of resonating sources, especially solids, and how the various properties of a resonating sound source (e.g., size, mass, tension) may affect sound source perception. In Chapter 3, Patterson, Smith, van Dinther, and Walters consider the standing-wave properties of sound sources, such as the vocal tract, and how the size of such resonators may determine the perception of the source. Chapter 4, by Demany and Semal, reviews much of the current knowledge about auditory memory, especially as it may relate to sound source perception. In addition to the importance of attending to one source or another to function in our everyday acoustic world, auditory attention may also play a direct role in aiding the auditory system in segregating one sound source from another. Chapter 5, by Hafter, Sarampalis, and Loui, reviews much of the literature related to auditory attention. In Chapter 6, by Kidd, Mason, Richards, Gallun, and Durlach, the topics of masking, especially energetic and informational masking, are reviewed as they relate to sound source perception. This is followed by Chapter 7, by Carylon and Gockel, in which the authors discuss how sources may be perceived and segregated based on a source's fundamental frequency of vibration and its resulting harmonic structure or temporal and spectral regularity.

A great deal is known about how differences in interaural arrival time and interaural level differences are used to locate the position of sound sources. Darwin (Chapter 8) considers the role spatial separation (especially interaural time and level differences) plays in sound source perception and segregation.

In Chapter 9, Sheft discusses temporal patterns of sounds and how these patterns pertain to sound source perception. This is followed by Chapter 10, by Lotto and Sullivan, who consider the speech-perception literature that provides insights into processes that might be considered for a better understanding of sound source perception for any potential sound source. Finally, in Chapter 11, Fay reviews the growing body of literature on how animal subjects other than humans process sound from sources.

Related chapters pertaining to other aspects of sound source perception can be found elsewhere in chapters from the Springer Handbook of Auditory Research series. Several chapters in Volume 15 (*Integrative Functions in the Mammalian Auditory Pathway*) relate to the questions of sound source perception, especially Chapter 9 ("Feature Detection in Auditory Cortex") by Nelken. Lewis and Fay, in Volume 22 (Chapter 2), *Evolution of the Vertebrate Auditory System*, is a treatment of the acoustic variables that could play a role in sound source perception. Volume 24 (*Pitch: Neural Coding and Perception*) contains several chapters that relate to this topic, especially Chapter 8, by Darwin, on "Pitch and Auditory Grouping." Volume 25 (*Sound Source Localization*) is a recent authoritative review of this topic. Volume 28 (*Hearing and Sound Communication in Amphibians*), and especially Chapter 11 on "Sound Processing in Real-World Environments," by Feng and Schul, treats many aspects of sound source perception in amphibians.

WILLIAM A. YOST, Chicago, IL
ARTHUR N. POPPER, College Park, MD
RICHARD R. FAY, Chicago, IL

Contributors

ROBERT P. CARLYON
MRC Cognition and Brain Sciences Unit, Cambridge CB2 7EF, UK, Email: bob.carlyon@mrc-cbu.cam.ac.uk

CHRISTOPHER J. DARWIN
Department of Psychology, School of Life Sciences, University of Sussex, Brighton BN1 9QG, UK, Email: cjd@sussex.ac.uk

LAURENT DEMANY
Laboratoire Mouvement, Adaptation, Cognition, CNRS and Université Victor Segalen, F-33076 Bordeaux, France, Email: laurent. demany@psyac.u-bordeaux2.fr

NATHANIEL I. DURLACH
Hearing Research Center, Boston University, Boston, MA, 02215, USA

RICHARD R. FAY
Parmly Hearing Institute, Loyola University Chicago, Chicago, IL, 60626, Email: rfay@luc.edu

FREDERICK J. GALLUN
Hearing Research Center, Boston University, Boston, MA, 02215, USA

HEDWIG E. GOCKEL
MRC Cognition and Brain Sciences Unit, Cambridge CB2 7EF, UK, Email: hedwig.gockel@mrc-cbu.cam.ac.uk

ERVIN R. HAFTER
Department of Psychology, University of California at Berkeley, Berkeley, CA, 94720, Email: hafter@berkeley.edu

GERALD KIDD, JR.
Hearing Research Center, Boston University, Boston, MA, 02215, USA, Email: gkidd@bu.edu

ANDREW J. LOTTO
Department of Speech, Language, and Hearing Sciences, University of Arizona, Tucson, AZ, 85721-0071, Email: alotto@email.arizona.edu

PSYCHE LOUI
Department of Psychology, University of California at Berkeley, Berkeley, CA, 94720, Email: psyche@berkeley.edu

ROBERT A. LUTFI
Auditory Behavioral Research Lab, Department of Communicative Disorders, University of Wisconsin – Madison, Madison, WI, 53706, Email: ralutfi@wisc.edu

CHRISTINE R. MASON
Hearing Research Center, Boston University, Boston, MA, 02215, USA

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

VIRGINIA M. RICHARDS
Hearing Research Center, Boston University, Boston, MA, 02215, USA

ANASTASIOS SARAMPALIS
Department of Psychology, University of California at Berkeley, Berkeley, CA, 94720, Email: asaram@berkeley.edu

CATHERINE SEMAL
Laboratoire Mouvement, Adaptation, Cognition, CNRS and Université Victor Segalen, F-33076 Bordeaux, France, Email: catherine. semal@psyac.u-bordeaux2.fr

STANLEY SHEFT
Parmly Hearing Institute, Loyola University Chicago, Chicago, IL, 60626, Email: ssheft@luc.edu

DAVID R.R. SMITH
Centre for the Neural Basis of Hearing, Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge CB2 3EG, UK, Email: drrs2@cam.ac.uk

SARAH C. SULLIVAN
Department of Psychology, University of Texas at Austin, Austin, TX 78712-0187, Email: sullivan@psy.utexas.edu

RALPH VAN DINTHER
Centre for the Neural Basis of Hearing, Department of Physiology, Development

and Neuroscience, University of Cambridge, Cambridge CB2 3EG, UK, Email: rv230@cam.ac.uk

THOMAS C. WALTERS
Centre for the Neural Basis of Hearing, Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge CB2 3EG, UK, Email: tcw24@cam.ac.uk

WILLIAM A. YOST
Speech and Hearing Sciences, Arizona State University, Tempe, AZ 85287, Email: William.yost@asu.edu

1
Perceiving Sound Sources

William A. Yost

1. Sound Source Perception

To survive, animals must navigate, find food, avoid predators, and reproduce; and many species survive based on their ability to communicate. All of these crucial behaviors allow animals to function in a crowded world of obstacles, objects, and other animals. Many of these objects vibrate and produce sound, which may be used to determine their sources and as a basis for communication. Thus, evolving an auditory system capable of processing sound provides animals a valuable ability to cope with the world.

Sounds produced by different sources are combined into one sound field that must be parsed into information that allows for the determination of the individual sources. The auditory peripheral receptors did not evolve to process the sources of sound, but instead, they provide the important neural code for the physical attributes of sound (i.e., of the sound field). While the neural code provides crucial information about the sounds from the sources that make up the sound field, the peripheral neural code is not a code for the sources themselves (but see Lewis and Fay 2004). Thus, the neural processes responsible for sound source determination lie above the auditory periphery. Given the variety of sources that produce sound, the complexity of the world in which these sources exist, and the lack of peripheral receptors to analyze sound sources per se, determining the sources of sound presents a significant challenge for the auditory system. At present, not a great deal is known about how the auditory system deals with this challenge. This book reviews several topics that are likely relevant to enhancing our understanding of the auditory system's ability to determine sound sources.

While not a great deal is known about sound source determination, research over the past century has produced a wealth of information about the neural processes involved with the coding and analysis of the physical attributes of sound and the sensations produced by these physical attributes. In the behavioral sciences, it is not uncommon, especially in vision research, to conceptualize sensory processing in terms of sensation and perception. Auditory sensation may be viewed as processing the physical attributes of sound: frequency, level, and time. Auditory perception may be the additional processing of those attributes that allow an organism to deal with the sources that produced the sound. Thus, determining the sources of sound may be viewed as a study of auditory perception.

The perception of "special sources," such as speech and music, may be subtopics of the general topic of the perception of sound sources. Thus, this book is mainly about auditory perception as it relates to sound source determination.

Most of the recent history of the study of hearing has been dominated by the investigation of sensory processing of the physical attributes of sound. This was not the case in the early nineteenth century, when the modern study of biology and behavioral science was emerging (see Boring 1942). The philosopher/scientist of those times viewed the direct investigation of humans' awareness of objects and events in the world as a way to resolve some of the conflicts of mind/body dualism. However, it was recognized that it was the physical attributes of the sound from sources that were being processed. Giants in science, such as von Helmholtz, developed theories for processing acoustic attributes, especially frequency. The investigation of the auditory processing of frequency, level, and time have dominated studies of hearing throughout the past 200 years, as a wealth of information was and is being obtained concerning auditory processing of the physical attributes of sound and its consequences for auditory sensations.

In the late 1980s and early 1990s, several authors wrote about object perception (Moore 1997), perception of auditory entities (Hartmann 1988), auditory images (Yost 1992), and sound source determination (Yost and Sheft 1993). Bregman's book, *Auditory Scene Analysis* (1990), provided a major contribution in restoring a significant interest in sound source perception. These works recognized that others previously had studied issues directly related to sound source perception. For instance, Cherry (1953) coined the term "cocktail party problem" as a way to conceptualize processing sound sources in complex multisource acoustic environments. Cherry proposed that several aspect of sound sources and auditory processing might help solve the "cocktail problem," and he investigated primarily one of these, binaural processing. Bregman's work leading up to *Auditory Scene Analysis* largely involved the study of auditory stream segregation. Thus, the topic of sound source perception has at times been referred to as object perception, image perception, entity perception, identification, the cocktail party problem, streaming, source segregation, sound source determination, and auditory scene analysis.

While at a general level these terms may be synonymous, in many contexts they may refer to different aspects of the study of sound source perception. Images, entities, and objects can refer to either the physical source or the perception of the source, and as such can present an ambiguous description of the problem. Identification implies using a label for the sound from a source. It is not clear that sound source perception requires the use of labels, i.e., that one has to be able to identify a sound in order for it to be processed as a source. Many procedures used to study sound source perception use identification, and we clearly identify (label) the sound from many sources, so that identification plays a role in sound source perception and its study. However, identification is probably not necessary and sufficient for sound source perception.

The cocktail party problem has often been identified with the ability to use binaural processing (or more generally spatial processing) to segregate sound sources. Auditory streaming usually refers to a context of alternating bursts of sound, which does not represent the sound from all sources (e.g., a single breaking twig). And many of these terms emphasize the segregation of sources rather than a more general problem of the perception of a sound source whether it produces sound in isolation or along with the sound from other sources. Thus, this book will attempt to use the term "sound source perception" as the general description of the challenge facing the auditory system.

Figure 1.1 depicts the type of neural information that might be provided to the central nervous system by the auditory periphery for a segment of a musical piece played by a quartet of a piano, bass, drum, and trumpet. In this segment all four instruments are playing together, and each can be identified when one listens to the piece. The figure is based on the computations of the Auditory Image Model (AIM; see Patterson et al. 1995), in which the sound is processed by a middle-ear transfer function, a simulation of basilar membrane analysis (a gammatone filter bank), and a simulation of hair cell and auditory nerve processing (Meddis hair cell). Each horizontal trace reflects a histogram of neural

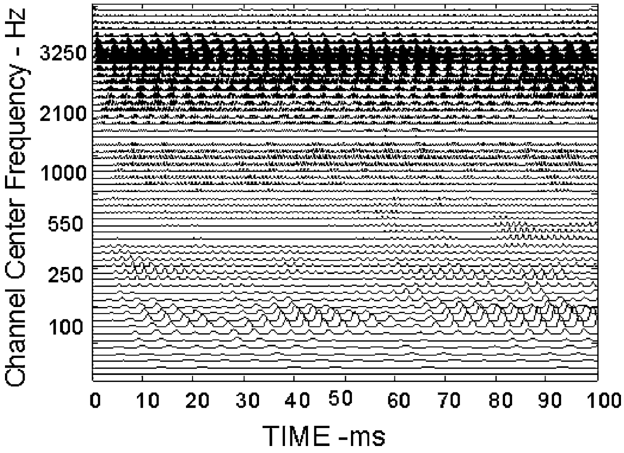

FIGURE 1.1. The neural activation pattern (NAP) for the simultaneous playing of a trumpet, bass, piano, and drum in a 100-ms slice of a piece of jazz music. The NAP represents an estimate of the pattern of neural information flowing from the auditory periphery (see Patterson et al. 1995). Each line in the NAP represents an estimate of the poststimulus time histogram of a small set of fibers tuned to a narrow frequency region indicated on the *y*-axis. Increased neural activity at different moments in time represents the neural code for temporal modulation, while increased neural activity for one set of tuned fibers over the others represents the neural code for spectral processing. The spectral-temporal pattern of each instrument overlaps that from the other instruments, and as such, the NAP represents the "summed" spectral-temporal pattern of the four instruments. The central nervous system must parse this overall spectral-temporal code into subsets each corresponding to an instrument (sound source) in order for the existence of each instrument in the musical piece to be determined. How this parsing is achieved is the challenge this book addresses.

activity from a small set of tuned fibers with the center frequency of each set of tuned fibers shown on the ordinate and time shown along the abscissa. While this depiction may not represent all details of the analysis performed by the auditory periphery, it does represent the challenge facing the central nervous system. This peripheral code indicates the spectral-temporal neural structure of the complex sound field consisting of the sounds from the four instruments, but it does not provide an obvious depiction of the four sources. That is, how would the central nervous system parse this spectral-temporal neural pattern into four subsets, each representing one of the four instruments?

Two approaches to evaluating the peripheral neural code have been used to account for sound source perception. Many investigators attempt to identify attributes of the sound from sources that are preserved in the neural code that could be used to segregate one source from another. Others have imagined that the problem of sound source perception is aided by the auditory system using information it has gained from experience. The former is a bottom-up approach, and the latter a top-down approach. Bregman (1990) argues that both are necessary to account for sound source perception, and he describes “primitive processes” as the bottom-up mechanisms and “schema-based approaches” as the top-down processes. That is, physical attributes such as temporal onset differences, harmonicity, and common modulation may be used by central nervous system circuits to segregate the neural code to account for sound source perception. At the same time, information gleaned from one’s experience in processing sound sources can guide these neural circuits in processing the peripheral neural code. Sheft, in Chapter 9, makes a distinction between segregating neural information to aid sound source perception and extracting information about the segregated neural information as an equally important part of perceiving sound sources.

2. Chapters in This Book

Sound sources produce sound based on their vibratory patterns, most often due to the resonance properties of the source. The vibration and resonances produce a sound pressure wave that is propagated from the source to a listener’s ears, and this sound wave is characterized by its frequency, level, and timing attributes. A great deal is known about how the auditory system processes these physical attributes. But do the properties of the source itself (e.g., size, mass, tension) have any direct bearing on auditory perception, and, if so, in what way? Chapters 2 (by Lutfi) and 3 (by Patterson, Smith, van Dinther, and Walters) review much of what is known about how the resonant properties of sound sources affect perception. Chapter 2 deals with how the various properties of a resonator, especially solid objects (e.g., a piece of wood that is struck) may affect sound source perception. Lutfi reviews how one characterizes the resonant properties of vibrating sources and several studies concerning the relationship between the physical aspects of resonance and one's auditory perceptions. The chapter introduces several newer psychophysical techniques used to study the perception of complex sounds.

Chapter 3 describes issues of the resonance properties of tubes (e.g., the vocal tract) that resonate due to standing waves. The size of a resonator often has significant biological relevance. In many species (e.g., frog), the female prefers to mate with large males and chooses mates based on the sounds produced by the male. Thus, males need to produce sounds that are correlated with size. In humans, the size of the vocal tract is correlated with the frequency content of speech (e.g., formants). A man and a woman each uttering the same vowel will produce different formant structures due largely to the difference in the size of the vocal tracts. Yet a listener still perceives the same vowel as being uttered by both people. How does the auditory system compensate for the formant differences in order that a constant percept is achieved? These are some of the topics covered in Chapter 3 (and to some extent in Chapter 10).

Sound is a temporal stimulus that is processed over time. The perception of a musical piece, a person speaking, a car driving past requires us to integrate the sound waveform over time. That is, sound that occurred a moment ago has to be retained by the auditory system so that the sound occurring now makes sense. The neural representation of sound must be continually stored and retrieved. How well the nervous system stores and retrieves its neural representations of sound will influence the perceptions of a sound source. The mechanisms of storage and retrieval may directly aid the processes of sound source determination. Researchers must often confront problems of a listener's ability to remember acoustic events when they investigate complex sound processing. Chapter 4, by Demany and Semal, reviews much of the current knowledge about auditory memory, especially as it may relate to sound source perception.

Many sound sources (people talking, music playing, glasses clinking, etc.) occur during a typical cocktail party. Often, the challenge is not just to segregate these sound sources, but to attend to one or more of them (e.g., to follow a friend's conversation). In addition to the importance of attending to one source or another to function in our everyday acoustic world, auditory attention may also play a direct role in aiding the auditory system in segregating one sound source from another. Chapter 5, by Hafter, Sarampalis, and Loui, reviews much of the literature related to auditory attention.

If one sound masks another, then it is unlikely that the masked sound will contribute much to sound source perception. Thus, masking clearly plays an important role in sound source perception, especially source segregation. Masking is typically defined as the elevation in detection threshold (but see Tanner 1958) of one sound (the signal) in the presence of another sound or other sounds (masker or maskers). In the 1980s, Watson and colleagues (see Watson 2005 for a review) showed that when listeners were presented an uncertain stimulus context, they had more difficulty processing sound than if the stimulus conditions were more certain or less variable; the greater the uncertainty, the greater the difficulty in sound processing. Borrowing from earlier work of Pollack (1975), Watson labeled the interference among different sounds that was based only on the sound itself “energetic masking.” The additional masking or interference due to making the stimulus context variable and uncertain was labeled “informational masking.” Carhart and colleagues (1969) were among the first to show that in studies of speech masking, a masker that was itself speech provided more masking for a speech signal than a nonspeech masker (e.g., noise) when the nonspeech and speech maskers are equated to the extent possible in terms of their acoustic properties. The additional masking that occurs for the speech masker over the nonspeech masker is now also referred to as informational masking, and the masking provided by the nonspeech masker is referred to as energetic masking. In the speech example, it has been suggested that the additional masking that occurs when both the target signal and inferring masker are speech is due to the auditory system’s attempt to segregate these two similar information-bearing sound sources. Thus, in the real world of complex sound sources, the challenge is to segregate sound sources that provide information for the listener and occur in uncertain contexts. Chapter 6, by Kidd, Mason, Richards, Gallun, and Durlach, reviews the topics of masking, especially energetic and informational masking, as they relate to sound source perception.

The auditory system has sophisticated biomechanical and neural processes to determine the spectral content of an acoustic event. Thus, it is logical to assume that aspects of a sound’s spectrum could be used in sound source perception, especially as an aid in segregating the sound from one source from that from other sources. In simple cases, the spectrum of the sound from one source may not overlap that from another source, and since the auditory system neurally codes for these spectral differences, this neural differentiation could form a basis for sound source segregation. But how does the auditory system segregate sound sources when the spectral content of one source overlaps that of other sources (as indicated in Figure 1.1)?

As Chapters 2 and 3 indicate, resonant sources often have a fundamental frequency with many harmonics. One source is likely to vibrate with a fundamental frequency that is different from that of another source. The vibratory pattern of a harmonic sound provides a temporal regularity to the waveform as well as spectral regularity. Thus, sources may be perceived and segregated based on the fundamental frequency and its resulting harmonic structure and/or on the resulting temporal and/or spectral regularity. Carylon and Gockel, in Chapter 7, review the current literature on the role harmonicity and regularity play in sound source perception. Often, sounds that are harmonic or contain either spectral or temporal regularity are perceived as having a distinct pitch (complex pitch). Sounds with different harmonic structures (or fundamentals) produce different pitches, so these pitch differences could be a basis for sound source perception. While it is still not certain how the auditory system extracts a pitch from such complex sounds (see Plack et al. 2006), it is clear that such processes involve using information across a broad range of the sound spectrum. Thus, studies of complex pitch also provide useful insights into how the nervous system performs spectral integration.

Cherry (1953) hypothesized that the fact that the source of a sound can be located in space may aid the auditory system in attending to one sound in the presence of sounds from other sources such as occurs at a cocktail party. That is, when different sources are located at different places, this spatial separation may allow the auditory system to segregate the sources. However, spatial separation is not both sufficient and necessary for sound source perception. The many instruments (sources) of an orchestra recorded by a single microphone and played over a single loudspeaker can be recognized; i.e., the sound sources (instruments) can be determined in the complete absence of spatial information.

A great deal is known about how differences in interaural arrival time and interaural level differences are used to locate the azimuth position of sound sources. Thus, it is not surprising that most investigations of spatial separation of sound sources have studied interaural time and level differences. Darwin, in Chapter 8, describes a great deal of the literature on the role of spatial separation (especially interaural time and level differences) for sound source perception and segregation.

Since sound has no spatial dimensions and the auditory periphery does not code for the spatial location of sound sources, the nervous system must "compute" spatial location. It does so based on the interaction of sound with objects in the path of the sound as it travels from its source to the middle ears. For instance, sound from a source off to one side of a listener interacts with the head such that the sound reaches one ear before the other and is less intense (due to the sound shadow caused by the head) at the one ear than at the other. The auditory system computes these interaural time and level differences as the basis for processing the azimuth location of the sound source. However, two sources at different locations that produce exactly the same sound at exactly the same time will not generate two different sets of interaural time and level differences. That is, spatial separation per se is unlikely to be a cue for sound source segregation. The sound from one source must be either spectrally or temporally different from that from another source before the auditory system can compute their spatial locations (e.g., compute different interaural time and level differences). Thus, it is likely that the use of spatial separation of sound sources as a cue for sound source segregation involves an interaction of spectral and/or temporal processing in combination with spatial computations performed by the auditory system.

As already alluded to in this chapter, the temporal pattern of the sound from one source is likely to be different from that of others. These temporal differences provide potential cues for sound source perception and segregation. The sound from one source is likely to start and stop at a different time from that of another source, and the amplitude modulation imparted to one sound by its source's physical properties is likely to differ from the amplitude modulation imparted by a source with different physical properties. These two temporal aspects of sounds—onset characteristics and amplitude modulation—have been extensively studied, and a great deal of this literature as it pertains to sound source perception is presented in Chapter 9, by Sheft.

Speech is clearly the sound of most importance to humans, at least in terms of communication. Speech is generated by the vocal cords and tract, and recognizing one speech sound as different from another requires processing differences in the vocal-tract source. Thus, speech recognition can be viewed as a case of sound source perception, although perhaps a special case. A great deal is known about the relationship between the sound source for speech, the vocal tract, and speech perception. Thus, one might learn about the general problem of sound source perception from studying speech perception. Lotto and Sullivan review in Chapter 10 several topics from the speech perception literature that provide insights into processes that could be considered for a better understanding of sound source perception for any potential sound source. It is clear from speech perception research that one's perception of a sound at one instance in time can be altered by another sound that occurred earlier and in some cases later in time. It is likely that such issues of coarticulation of speech sounds may also play a significant role in the perception of the sounds from other sources.

It is highly unlikely that sound source perception and segregation are unique to a limited number of animal species. It is hard to imagine how any animal with the ability to process sound would not need to deal with processing sound produced by different sources. Thus, sound source perception probably occurs for all animals. Fay, in Chapter 11, reviews some of the growing body of literature on how animal subjects other than humans process sound from sources. This research indicates how animals cope with the challenge of sound source perception, offers a chance to develop an understanding of universal processes that might be used by all animals as opposed to special processes used by only certain animals, and provides a deeper understanding of neural data related to sound source processing that are obtained from different animals. As Fay explains, little research has been devoted to studying sound source perception in animals other than humans. Much of that research uses the streaming paradigm developed by Bregman and colleagues (see Bregman 1990).

The topics covered in this book are likely to play a role in dealing with sound sources both in the laboratory and in the day-to-day world. These topics may play a direct role in how the auditory system processes sound from a source or multiple sources. For instance, everyday experience suggests that auditory attention is required to cope with most complex acoustic scenes. Attention appears to play an important role after the sound sources are processed (segregated), by allowing one to focus on one source as opposed to other sources. It is also possible that attentional mechanisms directly aid the auditory system in segregating sound sources in the first place. That is, attention may play a role in both the segregation process itself and in how the sources are dealt with after they are segregated.

Several topics covered in this book are important for conducting experiments dealing with sound source perception even if they may not always be directly related to general issues of sound source processing. For instance, in many experiments a sound at one instant in time must be compared to a sound occurring at another instant in time. This comparison is based on the earlier sound staying in some form of memory in order for it to be perceptually compared to the later sound. It is possible that the ability to process the sounds in such experiments is related not only to the physical properties of the two sounds and the context in which they exist, but also to memory.

3. Other Topics Related to Sound Source Perception

This book and its chapters are somewhat "theoretically neutral" in that the choice of chapters and the presentation in each chapter are not tightly coupled to a particular theory or even a particular theoretical framework. The theories of sound source perception and processing are few. As mentioned previously, Bregman (1990) offers a theoretical framework that helps him organize his views of processing the auditory scene. This framework is based partially on ideas shared with the Gestalt principles of perception. An ecological, or Gibsonian, approach has also been proposed (see Neuhoff 2004) in which sound source perception is viewed as arising from our experience in processing real sounds in the real world. There is also a growing literature on computational models of sound source processing, often referred to as computational auditory scene analysis (CASA); see Wang and Brown (2006).

In addition to not covering a great deal about theories and models of sound source perception, especially the growing use of CASA models, other topics are not covered in great detail in this book. For instance, there is a growing literature of physiological studies of sound source perception. While several chapters in this book review much of this literature, a single chapter was not devoted to physiological processes. The major reason these topics were not covered in detail was the desire to keep the length of the book consistent with the other books in the SHAR series.

But there were other reasons as well. Both the use of CASA models and physiological correlates to studying auditory source processing face a similar challenge. There are very few data on sound source perception per se. That is, there are few data directly related to the actual segregation of sound sources. Thus, it is difficult to provide models or to find physiological measures of actual data indicating how sound sources are segregated. The most common data set that is modeled and studied physiologically is that related to auditory streaming. While the studies of auditory streaming have provided valuable information pertaining to sound source perception, they represent only one class of stimulus conditions. The main streaming paradigm is to alternate two (or more) sounds in time, e.g., two tones of different frequency. Under the proper conditions the sounds are perceived as two pulsating sources (e.g., two pulsating tone bursts) as opposed to one source that alternates in some perceptual attribute (e.g., pitch). Many sound sources are modulated in level over time, and they are still perceived as a single source. Thus, understanding how the pulsating sound from one source provides a perception of a single source rather than