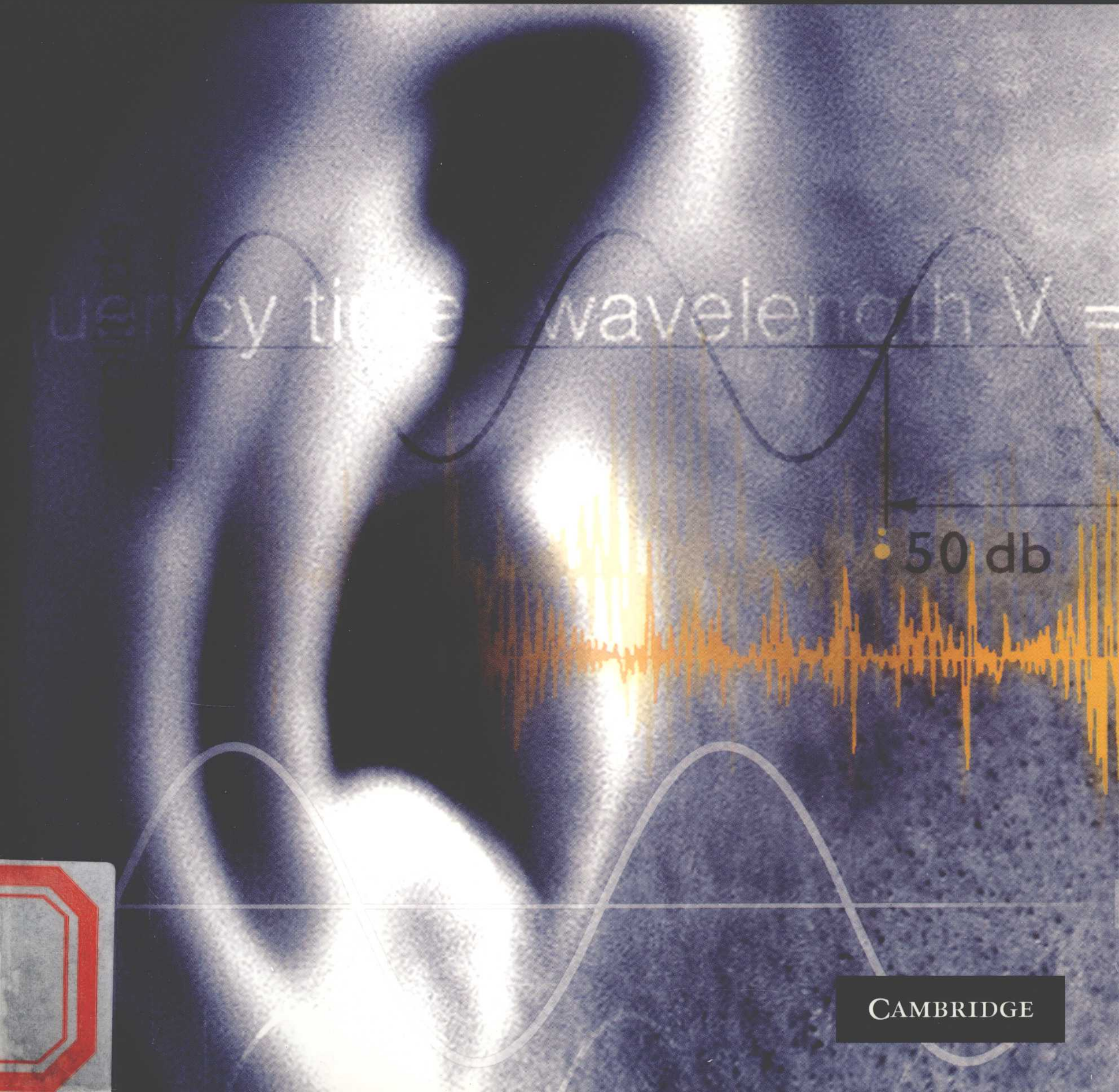
Richard M. Warren

Auditory Perception

THIRD EDITION

An Analysis and Synthesis



CAMBRIDGE

Auditory Perception

An Analysis and Synthesis

Third Edition

RICHARD M. WARREN

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo, Delhi

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521868709

First published 2008

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data
Warren, Richard M.
Auditory perception : an analysis and synthesis / Richard M. Warren. – 3rd ed.
p. ; cm.
Includes bibliographical references and index.
ISBN 978-0-521-86870-9 (hardback) – ISBN 978-0-521-68889-5 (pbk.) 1. Auditory perception. 2. Speech perception. I. Title.
[DNLM : 1. Auditory Perception. 2. Speech Perception. WV 272 W2831a 2008]
QP461.W27 2008
152.1′5–dc22 2007050033

ISBN 978-0-521-86870-9 hardback
ISBN 978-0-521-68889-5 paperback

Auditory Perception

An Analysis and Synthesis

This revised and updated Third Edition describes the nature of sound, how sound is analyzed by the auditory system, and the rules and principles governing our interpretation of auditory input. It covers many topics including sound and the auditory system, locating sound sources, the basis for loudness judgments, perception of acoustic sequences, perceptual restoration of obliterated sounds, speech production and perception, and the relation of hearing to perception in general. Whilst keeping the consistent style of the previous editions, many new features have been added, including suggestions for further reading at the end of each chapter, a section on functional imaging of the brain, expanded information on pitch and infrapitch, and additional coverage of speech processing. Advanced undergraduate and graduate students interested in auditory perception, behavioral sciences, psychology, neurobiology, architectural acoustics, and the hearing sciences will find this book an excellent guide.

RICHARD M. WARREN is Research Professor and Distinguished Professor Emeritus in the Department of Psychology at the University of Wisconsin-Milwaukee. He is a Fellow of the Acoustical Society of America, American Psychological Association, and the Association for Psychological Science.

Preface

As in the earlier editions, the present text emphasizes the interconnectedness of areas in auditory perception. These linkages are especially evident in the chapters dealing with acoustic sequences, pitch and infrapitch, loudness, and the restoration of portions of signals obliterated by extraneous sounds. In addition, the chapter on speech describes how processes employed for the perception of brief nonverbal sounds are used for the organization of syllables and words, along with an overlay of special linguistic mechanisms.

The basic format of the book remains unchanged, but all chapters have been updated. Among the additions are new sections in Chapter 1 describing the principles underlying functional imaging of the brain based on the hemodynamic techniques of fMRI and PET, and the electrodynamic techniques of EEG and MEG. New information concerning pitch and infrapitch appears in Chapter 3, and additional information concerning speech processing is incorporated into Chapter 7. Suggested additional reading now appears at the end of each chapter.

It is hoped that this text will be of value to research scientists and to professionals dealing with sound and hearing. No detailed specialized knowledge is assumed, since basic information necessary for understanding the material covered is provided. It may be used for advanced undergraduate and graduate courses in behavioral sciences, neurobiology, music, audio engineering, and the health sciences and professions.

My own research in perception was carried out at the following institutions: Brown University; New York University College of Medicine; Cambridge University; the Medical Research Council Applied Psychology Research Unit, Cambridge; Oxford University; the Laboratory of Psychology at the National Institute of Mental Health, Bethesda; and the University of Wisconsin-Milwaukee.

I acknowledge the debts to my graduate students over the years.

Dr. Peter W. Lenz has made essential contributions to all aspects of the research currently being carried out in our laboratory.

My debt to Jim Bashford is especially great: he has been my colleague and collaborator since the 1970s. Our back-and-forth discussions have played a basic role in designing and conducting the work in our laboratory.

I wish to thank Ms. Michelle L. Ullman for her valuable and thorough bibliographic work and in the preparation of the typescript.

I am grateful for the past research support by the National Research Council of the National Academy of Sciences, the National Science Foundation, the Air Force Office of Scientific Research, and the National Institutes of Health. My current support is from the National Institute on Deafness and Other Communication Disorders.

Finally, I acknowledge the essential role of Dr. Roslyn Pauker Warren, my colleague and wife. Without her, none of the editions of this book would have been started, and once started could not have been finished.

Please refer to www.cambridge.org/9780521868709 for audio demonstrations of some of the phenomena described in the text, that provide new insight into the mechanisms employed in auditory perception. The stimuli and descriptive narrative were produced by Dr. James A. Bashford, Jr.

Contents

1

Sound and the auditory system

This chapter provides a brief introduction to the physical nature of sound, the manner in which it is transmitted and transformed within the ear, and the nature of auditory neural responses.

The nature of auditory stimuli

The sounds responsible for hearing consist of rapid changes in air pressure that can be produced in a variety of ways – for example, by vibrations of objects such as the tines of a tuning fork or the wings of an insect, by puffs of air released by a siren or our vocal cords, and by the noisy turbulence of air escaping from a small opening. Sound travels through the air at sea level at a velocity of about 335 meters per second, or 1,100 feet per second, for all but very great amplitudes (extent of pressure changes) and for all waveforms (patterns of pressure changes over time). Special interest is attached to periodic sounds, or sounds having a fixed waveform repeated at a fixed frequency. Frequency is measured in hertz (Hz), or numbers of repetitions of a waveform per second; thus, 1,000 Hz corresponds to 1,000 repetitions of a particular waveform per second. The time required for one complete statement of an iterated waveform is its period. Periodic sounds from about 20 through 16,000 Hz can produce a sensation of pitch and are called tones. For reasons to be discussed shortly, it is generally considered that the simplest type of periodic sound is a sine wave or pure tone (shown in Figure 1.1A), which has a sinusoidal change in pressure over time. A limitless number of other periodic waveforms exists, including square waves (Figure 1.1B) and pulse trains (Figure 1.1C). Periodic sounds need not have simple, symmetrical waveforms: Figure 1.1D shows a periodic sound

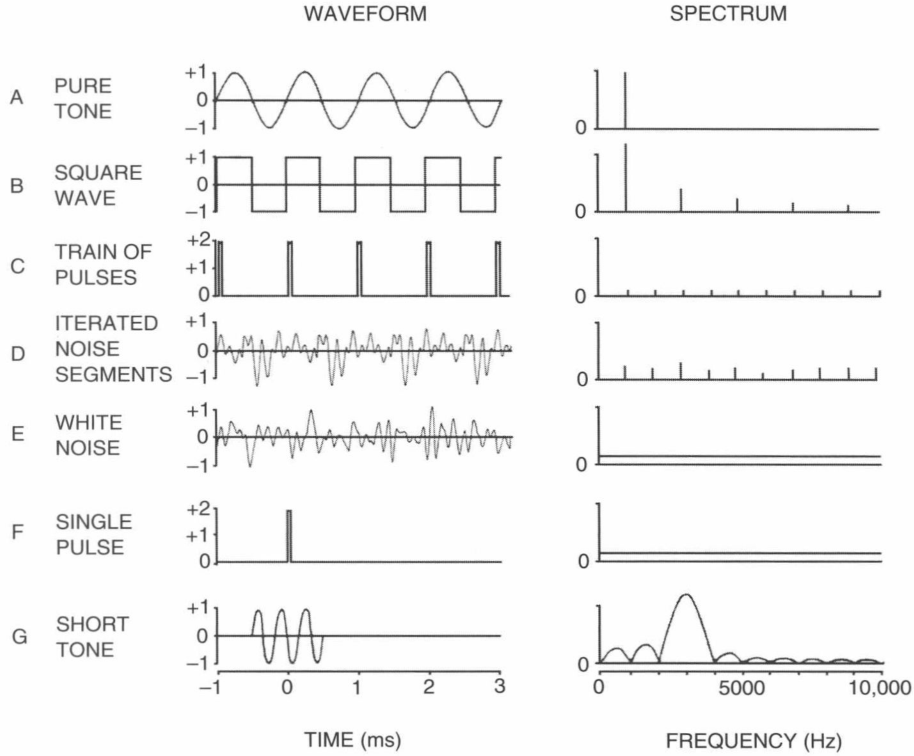


Figure 1.1 Waveforms and amplitude spectra. The waveforms A through E continue in time to produce the spectra as shown. Periodic waveforms A through D have line spectra, the others either continuous spectra (E and F), or a band spectrum (G). See the text for further discussion.

produced by iteration of a randomly generated waveform. The figure also depicts the waveforms of some nonperiodic sounds: white or Gaussian noise (Figure 1.1E), a single pulse (Figure 1.1F), and a short tone or tone burst (Figure 1.1G).

The waveforms shown in Figure 1.1 are time-domain representations in which both amplitude and time are depicted. Using a procedure developed by Joseph Fourier in the first half of the nineteenth century, one can represent any periodic sound in terms of a frequency-domain or spectral analysis in which a sound is described in terms of a harmonic sequence of sinusoidal components having appropriate frequency, amplitude, and phase relations. (Phase describes the portion of the period through which a waveform has advanced relative to an arbitrarily fixed reference time.) A sinusoidal or pure tone consists of a single spectral component, as shown in Figure 1.1A. The figure also shows the power spectra corresponding to the particular complex (nonsinusoidal) periodic sounds shown in Figures 1.1B, 1.1C, and 1.1D. Each of these sounds has a period of one millisecond, a fundamental frequency of 1,000 Hz (corresponding to the waveform repetition frequency), and harmonic components corresponding to integral multiples of the 1,000 Hz fundamental as indicated.

Frequency analysis is not restricted to periodic sounds: nonperiodic sounds also have a spectral composition as defined through use of a Fourier integral or Fourier transform (for details see Hartmann, 1998). Nonperiodic sounds have either continuous or band spectra rather than line spectra, as shown for the sounds depicted in Figures 1.1E, 1.1F, and 1.1G.

As we shall see, frequency analysis of both periodic and nonperiodic sounds is of particular importance in hearing, chiefly because a spectral analysis is performed within the ear leading to a selective stimulation of the auditory nerve fibers.

Although Figure 1.1 shows how particular waveforms can be analyzed in terms of spectral components, it is also possible to synthesize waveforms by adding together sinusoidal components of appropriate phase and amplitude. Figure 1.2 shows how a sawtooth waveform may be approximated closely by the mixing of only six harmonics having appropriate amplitude and phase.

The range of audible amplitude changes is very large. A sound producing discomfort may be as much as 10^6 times the amplitude level at threshold. Sound level can be measured as power as well as by amplitude or pressure at a particular point. Power usually can be considered as proportional to the square of the amplitude, so that discomfort occurs at a power level 10^{12} times the power threshold. The term "sound intensity" is, strictly speaking, the sound power arriving from a specified direction, and passing through a unit area perpendicular to that direction. However, the term "intensity" is often used interchangeably with "power," although the latter term has no directional specificity.

In order to span the large range of values needed to describe the levels of sound normally encountered, a logarithmic scale has been devised. The logarithm to the base 10 of the ratio of a particular sound power level to a reference power level defines the level of the sound in Bels (named in honor of Alexander Graham Bell). However, the Bel is a rather large unit, and it is conventional to use a unit 1/10 this size, the decibel (or dB) to express sound levels. The level in dB can be defined as:

$$\text{dB} = 10 \log_{10} (I_1/I_2)$$

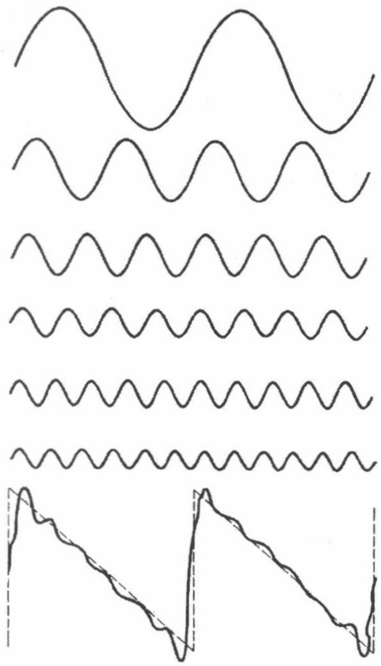


Figure 1.2 Synthesis of a complex waveform through addition of harmonically related sinusoidal components. The approximation of a sawtooth waveform could be made closer by the addition of higher harmonics of appropriate amplitude and phase. (From Brown and Deffenbacher, 1979.)

where I_1 is the intensity or power level of the particular sound of interest, and I_2 is the reference level expressed as sound intensity. One can also calculate decibels on the basis of pressure or amplitude units using the equation:

$$\mathrm{dB} = 20\log_{10}(P_1/P_2)$$

where P_1 is the relative pressure level being measured and P_2 is the reference pressure level. The standard reference pressure level is 0.0002 dyne/cm^2 (which is sometimes expressed in different units of 20 micropascals). The level in dB measured relative to this standard is called the Sound Pressure Level (or SPL). Sound-level meters are calibrated so that the numerical value of the SPL can be read out directly. There is another measure of sound level, also expressed in dB, called the Sensation Level (SL), which is used occasionally in psychoacoustics. When measuring SL, the level corresponding to the threshold of a sound for an individual listener is used as the reference level rather than the standard physical value employed for SPL, so that dB SL represents the level above an individual's threshold. Since SL is used relatively infrequently, dB will always refer to SPL unless otherwise specified.

To give some feeling for sound pressure levels in dB, the threshold of normal listeners for sinusoidal tones with frequencies between 1,000 and 4,000 Hz (the range exhibiting the lowest thresholds) is about 0 dB (the standard reference level); the ambient level (background noise) in radio and TV studios is about 30 dB, conversational speech about 65 dB, and the level inside a bus about 90 dB. Some rock bands achieve levels of 120 dB, which approaches the threshold for pain and can cause permanent damage to hearing following relatively brief exposures.

Experimenters can vary the relative intensities of spectral components by use of acoustic filters which, in analogy with light filters, pass only desired frequency components of a sound. A high-pass filter transmits only frequency components above a lower limit, and a low-pass filter transmits only frequencies below an upper limit. Bandpass filters (which transmit frequencies within a specified range) and band-reject filters (which block frequencies within a specified range) are available. The characteristics of high-pass and low-pass filters can be expressed in terms of both cut-off frequency (conventionally considered as the frequency at which the filter attenuation reduces power by half, or 3 dB), and the slope, or roll-off, which is usually expressed as dB/octave beyond the cut-off frequency (a decrease of one octave corresponds to halving the frequency). Bandpass filters are characterized by their bandwidth (the range in hertz between the upper and lower cut-off frequencies), and they can also be characterized by their "Q" (the bandwidth divided by the center frequency of the filter). In neurophysiological work, Q_{10} is sometimes used in which 10 dB downpoints are used to express the bandwidth rather than the conventional value of 3 dB. Filter types are shown in Figure 1.3.

Our auditory apparatus

The outer ear and the middle ear

It is convenient to consider the ear as consisting of three divisions. The outer ear, also called the pinna (plural "pinnae") or auricle, is shown in Figure 1.4. It appears to contribute to localization of sound sources by virtue of its direction-specific effect on the intensity of certain frequency components of sounds, as will be discussed in the next chapter. The human pinna is surrounded by a simple flange (the helix) which is extended considerably in some other mammals to form a conical structure functioning as a short version of the old-fashioned ear trumpet. These acoustic funnels can enhance sensitivity to high frequency sounds when pointed toward their source by controlling muscles, as well as being of help in locating the sound source.

After the acoustic transformation produced by reflections within our pinna, the sound passes through the ear canal (or external auditory meatus) which

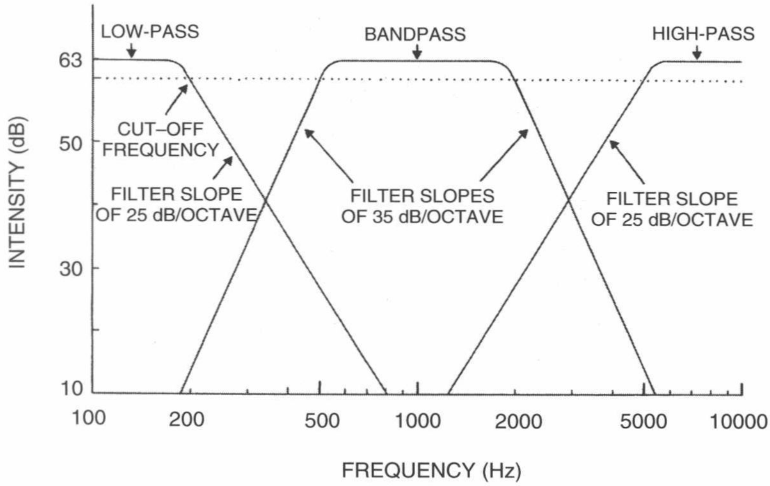


Figure 1.3 Characteristics of filters. Low-pass, high-pass, and bandpass filters are illustrated, along with filter slopes (dB/octave) and cut-off frequencies (frequencies at which there is a 3 dB reduction in intensity).

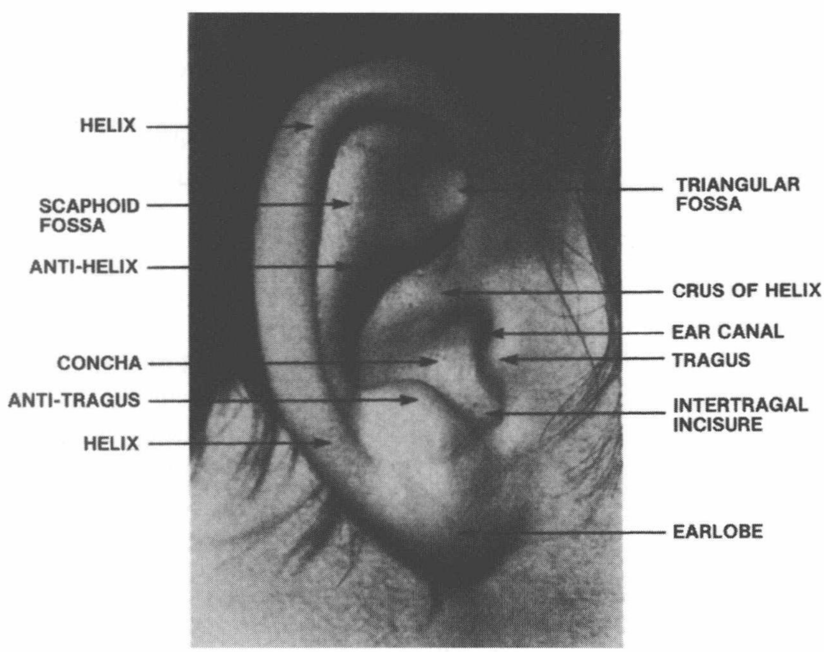


Figure 1.4 The outer ear (other names: pinna and auricle). The major anatomical features are shown.